PCT

INTERNATIONAL APPLICATION PUBLISHED UI

WO   9602561A1

(51) International Patent Classification 6 : C07H 21/04, C07K 19/00, C12N 1/19, 5/10, C12Q 1/68, G01N 33/68	A1	((43) International Publication Date: 1 February 1996 (01.02.96)

(21) International Application Number: PCT/US95/09307

(22) International Filing Date: 20 July 1995 (20.07.95)

(30) Priority Data:
08/278,082   20 July 1994 (20.07.94)   US

(71) Applicants: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, MA 02114 (US). GENETICS INSTITUTE, INC. [US/US]; 87 Cambridgepark Drive, Cambridge, MA 02140 (US).

(72) Inventors: BRENT, Roger; 27 Wendell Street, Cambridge, MA 02138 (US). MCCOY, John, M.; 63 Pine Ridge Road, Reading, MA 01867 (US). JESSEN, Timm, H.; Wiesen Weg 3, D-65812 Bad Soden (DE). XU, Chanxing, Wilson; Apartment 9P, 9 Hawthorne Place, Boston, MA 02114 (US).

(74) Agent: LECH, Karen, F.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States: JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: INTERACTION TRAP SYSTEMS FOR DETECTING PROTEIN INTERACTIONS

(57) Abstract

    Disclosed herein is a method of determining whether a first protein is capable of physically interacting with a second protein, involving: (a) providing a host cell which contains (i) a reporter gene operably linked to a protein binding site; (ii) a first fusion gene which expresses a first fusion protein, the first fusion protein including the first protein covalently bonded to a binding moiety which is capable of specifically binding to the protein binding site; and (iii) a second fusion gene which expresses a second fusion protein, the second fusion protein including the second protein covalently bonded to a gene activating moiety and being conformationally-constrained; and (b) measuring expression of the reporter gene as a measure of an interaction between the first and the second proteins. Also disclosed are methods for assaying protein interactions, and identifying antagonists and agonists of protein interactions. Proteins isolated by these methods are also discussed. Finally, populations of eukaryotic cells are disclosed, each cell having a recombinant DNA molecule encoding a conformationally-constrained intracellular peptide.

INTERACTION TRAP SYSTEMS FOR
DETECTING PROTEIN INTERACTIONS

Background of the Invention

5   This invention relates to methods for detecting
protein interactions and isolating novel proteins.

Summary of the Invention

In general, the invention features methods for
detecting interactions among proteins.

10   Accordingly, in one aspect, the invention features
a method of determining whether a first protein is
capable of physically interacting with a second protein.
The method includes (a) providing a host cell which
contains (i) a reporter gene operably linked to a DNA-

15 binding-protein recognition site; (ii) a first fusion
gene which expresses a first fusion protein, the first
fusion protein comprising the first protein covalently
bonded to a binding moiety which is capable of
specifically binding to the DNA-binding-protein

20 recognition site; and (iii) a second fusion gene which
expresses a second fusion protein, the second fusion
protein including the second protein covalently bonded to
a gene activating moiety and being conformationally-
constrained; and (b) measuring expression of the reporter

25 gene as a measure of an interaction between the first and
said second proteins.

Preferably, the second protein is a short peptide
of at least 6 amino acids in length and is less than or
equal to 60 amino acids in length; includes a randomly

30 generated or intentionally designed peptide sequence; or
is conformationally-constrained as a result of covalent
bonding to a conformation-constraining protein, e.g.,
thioredoxin or a thioredoxin-like molecule. Where the
second protein is covalently bonded to a conformationally

constraining protein the invention features a polypeptide wherein the second protein is embedded within the conformation-constraining protein to which it is covalently bonded. Where the conformation-constraining
5 protein is thioredoxin, the invention also features an additional method which includes a second protein which is conformationally-constrained by disulfide bonds between cysteine residues in the amino-terminus and in the carboxy-terminus of the second protein.

10 In another aspect, the invention features a method of detecting an interacting protein in a population of proteins, comprising: (a) providing a host cell which contains (i) a reporter gene operably linked to a DNA-binding-protein recognition site; and (ii) a fusion gene
15 which expresses a fusion protein, the fusion protein including a test protein covalently bonded to a binding moiety which is capable of specifically binding to the DNA-binding-protein recognition site; (b) introducing into the host cell a second fusion gene which expresses a
20 second fusion protein, the second fusion protein including one of said population of proteins covalently bonded to a gene activating moiety and being conformationally-constrained; and (c) measuring expression of the reporter gene. Preferably, the
25 population of proteins includes short peptides of between 1 and 60 amino acids in length.

The invention also features a method of detecting an interacting protein within a population wherein the population of proteins is a set of randomly generated or
30 intentionally designed peptide sequences, or where the population of proteins is conformationally-constrained by covalently bonding to a conformation-constraining protein. Preferably, where the population of proteins is conformationally-constrained by covalent bonding to a
35 conformation-constraining protein, the population of

- 3 -

proteins is embedded within the conformation-constraining
protein. The invention further features a method of
detecting an interacting protein within a population
wherein the conformation-constraining protein is
5 thioredoxin. Preferably, the population of proteins is
inserted into the active site loop of the thioredoxin.

The invention further features a method wherein
each of the population of proteins is conformationally-
constrained by disulfide bonds between cysteine residues
10 in the amino-terminus and in the carboxy-terminus of said
protein.

In preferred embodiments of various aspects, the
host cell is yeast; the DNA binding domain is LexA;
and/or the reporter gene is assayed by a color reaction
15 or by cell viability.

In other embodiments the bait may be Cdk2 or a Ras
protein sequence.

In another related aspect, the invention features
a method of identifying a candidate interactor. The
20 method includes (a) providing a reporter gene operably
linked to a DNA-binding-protein recognition site; (b)
providing a first fusion protein, which includes a first
protein covalently bonded to a binding moiety which is
capable of specifically binding to the DNA-binding-
25 protein recognition site; (c) providing a second fusion
protein, which includes a second protein covalently
bonded to a gene activating moiety and being
conformationally-constrained, the second protein being
capable of interacting with said first protein; (d)
30 contacting said candidate interactor with said first
protein and/or said second protein; and (e) measuring
expression of said reporter gene.

The invention features a method of identifying a
candidate interactor wherein the first fusion protein is
35 provided by providing a first fusion gene which expresses

- 4 -

the first fusion protein and wherein the second fusion
protein is provided by providing a second fusion gene
which expresses said second fusion protein.
(Alternatively, the reporter gene, the first fusion gene,
5   and the second fusion gene are included on a single piece
of DNA.)

        The invention also features a method of
identifying candidate interactors wherein the first
fusion protein and the second fusion protein are
10 permitted to interact prior to contact with said
candidate interactor, and a related method wherein the
first fusion protein and the candidate interactor are
permitted to interact prior to contact with said second
fusion protein.

15   In a preferred embodiment, the candidate
interactor is conformationally-constrained. Where the
candidate interactor is an antagonist, reporter gene
expression is reduced. Where the candidate interactor is
an agonist, reporter gene expression is increased. The
20 candidate interactor is a member selected from the group
consisting of proteins, polynucleotides, and small
molecules. In addition, a candidate interactor can be
encoded by a member of a cDNA or synthetic DNA library.
Moreover, the candidate interactor can be a mutated form
25 of said first fusion protein or said second fusion
protein.

        In a related aspect, the invention features a
population of eukaryotic cells, each cell having a
recombinant DNA molecule encoding a conformationally-
30 constrained intracellular peptide, there being at least
100 different recombinant molecules in the population,
each molecule being in at least one cell of said
population.

        Preferably, the intracellular peptides within the
35 population of cells are conformationally-constrained

- 5 -

because they are covalently bonded to a conformation-
constraining protein.

In preferred embodiments the intracellular peptide
is embedded within the conformation-constraining protein,
5   preferably thioredoxin; the intracellular peptide is
conformationally-constrained by disulfide bonds between
cysteine residues in the amino-terminus and in the
carboxy-terminus of said second protein; the population
of eukaryotic cells are yeast cells; the recombinant DNA
10 molecule further encodes a gene activating moiety
covalently bonded to said intracellular peptide; and/or
the intracellular peptide physically interacts with a
second recombinant protein inside said eukaryotic cells.

In another aspect, the invention features a method
15 of assaying an interaction between a first protein and a
second protein. The method includes: (a) providing a
reporter gene operably linked to a DNA-binding-protein
recognition site; (b) providing a first fusion protein
including a first protein covalently bonded to a binding
20 moiety which is capable of specifically binding to the
DNA-binding-protein recognition site; (c) providing a
second fusion protein including a second protein which is
conformationally constrained and covalently bonded to a
gene activating moiety; (d) combining the reporter gene,
25 the first fusion protein, and the second fusion protein;
and (e) measuring expression of the reporter gene.

The invention further features a method of
assaying the interaction between two proteins wherein the
first fusion protein is provided by providing a first
30 fusion gene which expresses the first fusion protein and
wherein the second fusion protein is provided by
providing a second fusion gene which expresses the second
fusion protein.

In yet other aspect, the invention features a
35 protein including the sequence Leu-Val-Cys-Lys-Ser-Tyr-

- 6 -

Arg-Leu-Asp-Trp-Glu-Ala-Gly-Ala-Leu-Phe-Arg-Ser-Leu-Phe
(SEQ ID NO: 1), preferably conformationally-constrained;
protein including the sequence Met-Val-Val-Ala-Ala-Glu-
Ala-Val-Arg-Thr-Val-Leu-Leu-Ala-Asp-Gly-Gly-Asp-Val-Thr
5 (SEQ ID NO: 2); preferably conformationally-constrained;
a protein including the sequence Pro-Asn-Trp-Pro-His-Gln-
Leu-Arg-Val-Gly-Arg-Val-Leu-Trp-Glu-Arg-Leu-Ser-Phe-Glu
(SEQ ID NO: 3), preferably conformationally-constrained;
a protein including the sequence Ser-Val-Arg-Met-Arg-Tyr-
10 Gly-Ile-Asp-Ala-Phe-Phe-Asp-Leu-Gly-Gly-Leu-Leu-His-Gly
(SEQ ID NO: 9), preferably conformationally-constrained;
a protein including the sequence Glu-Leu-Arg-His-Arg-Leu-
Gly-Arg-Ala-Leu-Ser-Glu-Asp-Met-Val-Arg-Gly-Leu-Ala-Trp-
Gly-Pro-Thr-Ser-His-Cys-Ala-Thr-Val-Pro-Gly-Thr-Ser-Asp-
15 Leu-Trp-Arg-Val-Ile-Arg-Phe-Leu (SEQ ID NO: 10),
preferably conformationally-constrained; a protein
including the sequence Tyr-Ser-Phe-Val-His-His-Gly-Phe-
Phe-Asn-Phe-Arg-Val-Ser-Trp-Arg-Glu-Met-Leu-Ala (SEQ ID
NO: 11), preferably conformationally-constrained; a
20 protein including the sequence Gln-Val-Trp-Ser-Leu-Trp-
Ala-Leu-Gly-Trp-Arg-Trp-Leu-Arg-Arg-Tyr-Gly-Trp-Asn-Met
(SEQ ID NO: 12), preferably conformationally-constrained;
a protein including the sequence Trp-Arg-Arg-Met-Glu-Leu-
Asp-Ala-Glu-Ile-Arg-Trp-Val-Lys-Pro-Ile-Ser-Pro-Leu-Glu
25 (SEQ ID NO: 13), preferably conformationally-constrained;
a protein including the sequence Trp-Ala-Glu-Trp-Cys-Gly-
Pro-Val-Cys-Ala-His-Gly-Ser-Arg-Ser-Leu-Thr-Leu-Leu-Thr-
Lys-Tyr-His-Val-Ser-Phe-Leu-Gly-Pro-Cys-Lys-Met-Ile-Ala-
Pro-Ile-Leu-Asp (SEQ ID NO:17), preferably
30 conformationally-constrained; a protein including the
sequence Leu-Val-Cys-Lys-Ser-Tyr-Arg-Leu-Asp-Trp-Glu-Ala-
Gly-Ala-Leu-Phe-Arg-Ser-Leu-Phe (SEQ ID NO: 18),
preferably conformationally-constrained; a protein
including the sequence Tyr-Arg-Trp-Gln-Gln-Gly-Val-Val-
35 Pro-Ser-Asn-Trp-Ala-Ser-Cys-Ser-Phe-Arg-Cys-Gly (SEQ ID

- 7 -

NO: 19), preferably conformationally-constrained; a
protein including the sequence Ser-Ser-Phe-Ser-Leu-Trp-
Leu-Leu-Met-Val-Lys-Ser-Ile-Lys-Arg-Ala-Ala-Trp-Glu-Leu-
Gly-Pro-Ser-Ser-Ala-Trp-Asn-Thr-Ser-Gly-Trp-Ala-Ser-Leu-
5   Ala-Asp-Phe-Tyr (SEQ ID NO: 20) preferably
conformationally-constrained; and substantially pure DNA
encoding the immediately foregoing proteins.

        The invention also includes novel proteins and
other candidate interactors identified by the foregoing
10 methods. It will be appreciated that these proteins and
candidate interactors may either increase or decrease
reporter gene activity and that these changes in activity
may be measured using assays described herein or known in
the art.

15   As used herein, by "reporter gene" is meant a gene
whose expression may be assayed; such genes include,
without limitation, lacZ, amino acid biosynthetic genes,
e.g. the yeast LEU2, HIS3, LYS2, TRP1, or URA3 genes,
nucleic acid biosynthetic genes, the mammalian
20 chloramphenicol transacetylase (CAT) gene, or any surface
antigen gene for which specific antibodies are available.
Reporter genes may encode any protein that provides a
phenotypic marker, for example, a protein that is
necessary for cell growth or a toxic protein leading to
25 cell death, or may encode a protein detectable by a color
assay leading to the presence or absence of color (e.g.,
florescent proteins and derivatives thereof).
Alternatively, a reporter gene may encode a suppressor
tRNA, the expression of which produces a phenotype that
30 can be assayed. A reporter gene according to the
invention includes elements (e.g., all promoter elements)
necessary for reporter gene function.

        By "operably linked" is meant that a gene and a
regulatory sequence(s) are connected in such a way as to
35 permit gene expression when the appropriate molecules

- 8 -

(e.g., transcriptional activator proteins or proteins which include transcriptional activation domains) are bound to the regulatory sequence(s).

By "covalently bonded" is meant that two domains
5 are joined by covalent bonds, directly or indirectly. That is, the "covalently bonded" proteins or protein moieties may be immediately contiguous or may be separated by stretches of one or more amino acids within the same fusion protein.

10   By "providing" is meant introducing the fusion proteins into the interaction system sequentially or simultaneously, and directly (as proteins) or indirectly (as genes encoding those proteins).

By "protein" is meant a sequence of amino acids of
15 any length, constituting all or a part of a naturally-occurring polypeptide or peptide, or constituting a non-naturally-occurring polypeptide or peptide (e.g., a randomly generated peptide sequence or one of an intentionally designed collection of peptide sequences).

20   By a "binding moiety" is meant a stretch of amino acids which is capable of directing specific polypeptide binding to a particular DNA sequence (i.e., a "DNA-binding-protein recognition site").

By "weak gene activating moiety" is meant a
25 stretch of amino acids which is capable of weakly inducing the expression of a gene to whose control region it is bound. As used herein, "weakly" is meant below the level of activation effected by GAL4 activation region II (Ma and Ptashne, Cell 48:847, 1987) and is preferably at
30 or below the level of activation effected by the B112 activation domain of Ma and Ptashne (Cell 51:113, 1987). Levels of activation may be measured using any downstream reporter gene system and comparing, in parallel assays, the level of expression stimulated by the GAL4 region II-

- 9 -

polypeptide with the level of expression stimulated by
the polypeptide to be tested.

By "altering the expression of the reporter gene"
is meant an increase or decrease in the expression of the
5 reporter gene to the extent required for detection of a
change in the assay being employed. It will be
appreciated that the degree of change will vary depending
upon the type of reporter gene construct or reporter gene
expression assay being employed.
10   By "conformationally-constrained" is meant a
protein that has reduced structural flexibility because
its amino and carboxy termini are fixed in space.
Preferably, the conformationally-constrained protein is
displayed in a structurally rigid manner. Conformational
15 constraint according to the invention may be brought
about by exploiting the disulfide-bonding ability of a
natural or recombinantly-introduced pair of cysteine
residues, one residing at or near the amino-terminal end
of the protein of interest and the other at or near the
20 carboxy-terminal end. Alternatively, conformational
constraint may be facilitated by embedding the protein of
interest within a conformation-constraining protein.

By "conformation-constraining protein" is meant
any peptide or polypeptide which is capable of reducing
25 the flexibility of another protein's amino and/or carboxy
termini. Preferably, such proteins provide a rigid
scaffold or platform for the protein of interest. In
addition, such proteins preferably are capable of
providing protection·from proteolytic degradation and the
30 like, and/or are capable of enhancing solubility.
Examples of conformation-constraining proteins include
thioredoxin and other thioredoxin-like proteins,
nucleases (e.g., RNase A), proteases (e.g., trypsin),
protease inhibitors (e.g., bovine pancreatic trypsin
35 inhibitor), antibodies or structurally-rigid fragments

thereof, and conotoxins. A conformation-constraining
peptide can be of any appropriate length and can even be
a single amino acid residue.

        "Thioredoxin-like proteins" are defined herein as
5   amino acid sequences substantially similar, e.g., having
at least 18% homology, with the amino acid sequence of E.
coli thioredoxin over an amino acid sequence length of 80
amino acids. Alternatively, a thioredoxin-like DNA
sequence is defined herein as a DNA sequence encoding a
10 protein or fragment of a protein characterized by having
a three dimensional structure substantially similar to
that of human or E. coli thioredoxin, e.g., glutaredoxin
and optionally by containing an active-site loop. The
DNA sequence of glutaredoxin is an example of a
15 thioredoxin-like DNA sequence which encodes a protein
that exhibits such substantial similarity in three-
dimensional conformation and contains a Cys....Cys
active-site loop. The amino acid sequence of E. coli
thioredoxin is described in Eklund et al., EMBO J.
20 3:1443-1449 (1984). The three-dimensional structure of
E. coli thioredoxin is depicted in Fig. 2 of Holmgren, J.
Biol. Chem. 264:13963-13966 (1989). A DNA sequence
encoding the E. coli thioredoxin protein is set forth in
Lim et al., J. Bacteriol., 163:311-316 (1985). The three
25 dimensional structure of human thioredoxin is described
in Forman-Kay et al., Biochemistry 30:2685-98 (1991). A
comparison of the three dimensional structures of E. coli
thioredoxin and glutaredoxin is published in Xia, Protein
Science I:310-321 (1992). These four·publications are
30 incorporated herein by reference for the purpose of
providing information on thioredoxin-like proteins that
is known to one of skill in the art. Examples of
thioredoxin-like proteins are described herein.

        By "candidate interactors" is meant proteins
35 ("candidate interacting proteins") or compounds which

- 11 -

physically interact with a protein of interest; this term
also encompasses agonists and antagonists. Agonist
interactors are identified as compounds or proteins that
have the ability to increase reporter gene expression
5   mediated by a pair of interacting proteins. Antagonist
interactors are identified as compounds or proteins that
have the ability to decrease reporter gene expression
mediated by a pair of interacting proteins.

        "Compounds" include small molecules, generally
10   under 1000 MW, carbohydrates, polynucleotides, lipids,
and the like.

        By "test protein" is meant one of a pair of
interacting proteins, the other member of the pair
generally referred to as a "candidate interactor"
15   (supra).

        By "randomly generated" is meant sequences having
no predetermined sequence; this is contrasted with
"intentionally designed" sequences which have a DNA or
protein sequence or motif determined prior to their
20   synthesis.

        By "mutated" is meant altered in sequence, either
by site-directed or random mutagenesis. A mutated form
of a protein encompasses point mutations as well as
insertions, deletions, or rearrangements.
25   By "intracellular" is meant that the peptide is
localized inside the cell, rather than on the cell
surface.

        By an "activated Ras" is meant any mutated form of
Ras which remains bound to GTP for a period of time
30   longer than that exhibited by the corresponding wild-type
form of the protein. By "Ras" is meant any form of Ras
protein including, without limitation, N-ras, K-ras, and
H-ras.

        The interaction trap systems described herein
35   provide advantages over more conventional methods for

- 12 -

isolating interacting proteins or genes encoding
interacting proteins. For example, applicants' systems
provide rapid and inexpensive methods having very general
utility for identifying and purifying genes encoding a
5 wide range of useful proteins based on the protein's
physical interaction with a second polypeptide. This
general utility derives in part from the fact that the
components of the systems can be readily modified to
facilitate detection of protein interactions of widely
10 varying affinity (e.g., by using reporter genes which
differ quantitatively in their sensitivity to a protein
interaction). The inducible nature of the promoter used
to express the interacting proteins also increases the
scope of candidate interactors which may be detected
15 since even proteins whose chronic expression is toxic to
the host cell may be isolated simply by inducing a short
burst of the protein's expression and testing for its
ability to interact and stimulate expression of a
reporter gene.
20   If desired, detection of interacting proteins may
be accomplished through the use of weak gene activation
domain tags. This approach avoids restrictions on the
pool of available candidate interacting proteins which
may be associated with stronger activation domains (such
25 as GAL4 or VP16); although the mechanism is unclear, such
a restriction apparently results from low to moderate
levels of host cell toxicity mediated by the strong
activation domain.
         In addition, the claimed methods make use of
30 conformationally-constrained proteins (i.e., proteins
with reduced flexibility due to constraints at their
amino and carboxy termini). Conformational constraint
may be brought about by embedding the protein of interest
within a conformation-constraining protein (i.e., a
35 protein of appropriate length and amino acid composition

- 13 -

to be capable of locking the candidate interacting
protein into a particular three-dimensional structure).
Examples of conformation-constraining proteins include,
but are not limited to, thioredoxin (or other
5 thioredoxin-like proteins), nucleases (e.g., RNase A),
proteases (e.g., trypsin), protease inhibitors (e.g.,
bovine pancreatic trypsin inhibitor), antibodies or
structurally-rigid fragments thereof, and conotoxins.
        Alternatively, conformational constraint may be
10 accomplished by exploiting the disulfide-bonding ability
of a natural or recombinantly-introduced pair of cysteine
residues, one residing at the amino terminus of the
protein of interest and the other at its carboxy
terminus. Such disulfide bonding locks the protein into
15 a rigid and therefore conformationally-constrained loop
structure. Disulfide bonds between amino-terminal and
carboxy-terminal cysteines may be formed, for example, in
the cytoplasm of E. coli trxB mutant strains. Under
some conditions disulfide bonds may also form within the
20 cytoplasm and nucleus of higher organisms harboring
equivalent mutations, for example, an S. cerevisiae YTR4⁻
mutant strain (Furter et al., Nucl Acids Res. 14:6357-
6373, 1986; GenBank Accession Number P29509). In
addition, the thioredoxin fusions described herein (trxA
25 fusions) are amenable to this alternative means of
introducing conformational constraint, since the
cysteines at the base of peptides inserted within the
thioredoxin active-site loop are at a proper distance
from one another to form disulfide bonds under
30 appropriate conditions.
        Conformationally-constrained proteins as candidate
interactors are useful in the invention because they are
amenable to tertiary structural analysis, thus
facilitating the design of simple organic molecule
35 mimetics with improved pharmacological properties. For

- 14 -

example, because thioredoxin has a known structure, the
protein structure between the conformationally
constrained regions may be more easily solved using
methods such as NMR and X-ray difference analysis.
5 Certain conformation-constraining proteins also protect
the embedded protein from cellular degradation and/or
increase the protein's solubility, and/or otherwise alter
the capacity of the candidate interactor to interact.
        Once isolated, interacting proteins can also be
10 analyzed using the interaction trap system, with the
signal generated by the interaction being an indication
of any change in the proteins' interaction capabilities.
In one particular example, an alteration is made (e.g.,
by standard in vivo or in vitro directed or random
15 mutagenesis procedures) to one or both of the interacting
proteins, and the effect of the alteration(s) is
monitored by measuring reporter gene expression. Using
this technique, interacting proteins with increased or
decreased interaction potential are isolated. Such
20 proteins are useful as therapeutic molecules (for
example, agonists or antagonists) or, as described above,
as models for the design of simple organic molecule
mimetics.
        Protein agonists and antagonists may also be
25 readily identified and isolated using a variation of the
interaction trap system. In particular, once a protein-
protein interaction has been recorded, an additional DNA
coding for a candidate agonist or antagonist, or
preferably, one of a library of potential agonist- or
30 antagonist-encoding sequences is introduced into the host
cell, and reporter gene expression is measured.
Alternatively, candidate interactor agonist or antagonist
compounds (i.e., including polypeptides as well as non-
proteinaceous compounds, e.g., single stranded
35 polynucleotides) are introduced into an in vivo or in

- 15 -

<u>vitro</u> interaction trap system according to the invention
and their ability to effect reporter gene expression is
measured. A decrease in reporter gene expression
(compared to a control lacking the candidate sequence or
5 compound) indicates an antagonist. Conversely, an
increase in reporter gene expression (compared again to a
control) indicates an agonist. Interaction agonists and
antagonists are useful as therapeutic agents or as models
to design simple mimetics; if desired, an agonist or
10 antagonist protein may be conformationally-constrained to
provide the advantages described herein. Particular
examples of interacting proteins for which antagonists or
agonists may be identified include, but are not limited
to, the IL-6 receptor-ligand pair, TGF-β receptor-ligand
15 pair, IL-1 receptor-ligand pair and other receptor-ligand
interactions, protein kinase-substrate pairs, interacting
pairs of transcription factors, interacting components of
signal transduction pathways (for example, cytoplasmic
domains of certain receptors and G-proteins), pairs of
20 interacting proteins involved in cell cycle regulation
(for example, p16 and CDK4), and neurotransmitter pairs.
        Also included in the present invention are
libraries encoding conformationally-constrained proteins.
Such libraries (which may include natural as well as
25 synthetic DNA sequence collections) are expressed
intracellularly or, optionally, in cell-free systems, and
may be used together with any standard genetic selection
or screen or with any of a number of interaction trap
formats for the identification of interacting proteins,
30 agonist or antagonist proteins, or proteins that endow a
cell with any identifiable characteristic, for example,
proteins that perturb cell cycle progression.
Accordingly, peptide-encoding libraries (either random or
designed) can be used in selections or screens which
35 either are or are not transcriptionally-based. These

- 16 -

libraries (which preferably include at least 100
different peptide-encoding species and more preferably
include 1000, or 100,000 or greater individual species)
may be transformed into any useful prokaryotic or
5 eukaryotic host, with yeast representing the preferred
host. Alternatively, such peptide-encoding libraries may
be expressed in cell-free systems.

　　　Other features and advantages of the invention
will be apparent from the following detailed description
10 thereof, and from the claims.

<u>Brief Description of the Drawings</u>

　　　The drawings are first briefly described.

　　　FIGS. 1A-1C illustrate one interaction trap system
according to the invention.

15 　　　FIG. 2 is a diagram of a library vector pJM1.

　　　FIG. 3A is a photograph showing the interaction of
peptide aptamers with other proteins.

　　　FIG. 3B illustrates the sequence of exemplary Cdk2
interacting peptides.

20 　　　FIG. 4 illustrates coprecipitation of peptides 3
and 13 by Gst-Cdk2. Lane 1. Gst Beads, extract contains
TrxA;

Lane 2. Gst Beads, extract contains TrxA-peptide 3;

Lane 3. Gst Beads, extract contain TrxA-peptide 13;

25 Lane 4. Gst-Cdk2 beads, extract contains TrxA;

Lane 5. Gst-Cdk2 beads, extract contains TrxA-peptide 3;
and

Lane 6. Gst-Cdk2, extract contains TrxA-peptide 13.

　　　FIG. 5 illustrates the vector BRM116-H-Ras(G12V).

30 　　　FIG. 6 illustrates the vector pEG202-H-Ras(G12V).

<u>Detailed Description</u>

　　　Applicants have developed a novel interaction trap
system for the identification and analysis of
conformationally-constrained proteins that either
35 physically interact with a second protein of interest or

- 17 -

that antagonize or agonize such an interaction. In one
embodiment, the system involves a eukaryotic host strain
(e.g., a yeast strain) which is engineered to produce a
protein of therapeutic or diagnostic interest as a fusion
5   protein covalently bonded to a known DNA binding domain;
this protein is referred to as a "bait" protein because
its purpose in the system is to "catch" useful, but as
yet unknown or uncharacterized, interacting polypeptides
(termed the "prey"; see below). The eukaryotic host
10   strain also contains one or more "reporter genes," i.e.,
genes whose transcription is detected in response to a
bait-prey interaction. Bait proteins, via their DNA
binding domain, bind to their specific DNA recognition
site upstream of a reporter gene; reporter transcription
15   is not stimulated, however, because the bait protein
lacks an activation domain.
        To isolate DNA sequences encoding novel
interacting proteins, members of a DNA expression library
(e.g., a cDNA or synthetic DNA library, either random or
20   intentionally biased) are introduced into the strain
containing the reporter gene and bait protein; each
member of the library directs the synthesis of a
candidate interacting protein fused to an invariant gene
activation domain tag. Those library-encoded proteins
25   that physically interact with the promoter-bound bait
protein are referred to as "prey" proteins. Such bound
prey proteins (via their activation domain tag)
detectably activate the expression of the downstream
reporter gene and provide a ready assay for identifying a
30   particular DNA clone encoding an interacting protein of
interest. In the instant invention, each candidate prey
protein is conformationally-constrained (for example,
either by embedding the protein within a conformation-
constraining protein or by linking together the protein's
35   amino and carboxy termini). Such a protein is maintained

- 18 -

in a fixed, three-dimensional structure, facilitating
mimetic drug design.

An example of one interaction trap system
according to the invention is shown in Figures 1 A-C.
5   Figure 1A shows a leucine auxotroph yeast strain
containing two reporter genes, LexAop-LEU2 and LexAop-
lacZ, and a constitutively expressed bait protein gene.
The bait protein (shown as a pentagon) is fused to a DNA
binding domain (shown as a circle). The DNA binding
10   protein recognizes and binds a specific DNA-binding-
protein recognition site (shown as a solid rectangle)
operably-linked to a reporter gene. In Figures 1B and
1C, the cells additionally contain candidate prey
proteins (candidate interactors) (shown as an empty
15   rectangle in 1B and an empty hexagon in 1C) fused to an
activation domain (shown as a solid square); each prey
protein is embedded in a conformation-constraining
protein (shown as two solid half circles). Figure 1B
shows that if the candidate prey protein does not
20   interact with the transcriptionally-inert LexA-fusion
bait protein, the reporter genes are not transcribed; the
cell cannot grow into a colony on leu⁻ medium, and it is
white on Xgal medium because it contains no β-
galactosidase activity. Figure 1C shows that, if the
25   candidate prey protein interacts with the bait, both
reporter genes are active; the cell forms a colony on
leu⁻ medium, and cells in that colony have β-
galactosidase activity and are blue on Xgal medium.
Preferably, in this system, the bait protein (i.e., the
30   protein containing a site-specific DNA binding domain) is
transcriptionally inert, and the reporter genes (which
are bound by the bait protein) have essentially no basal
transcription.

Each component of the system is now described in
35   more detail.

Bait Proteins

      The selection host strain depicted in Figures 1 A-
C contains a DNA encoding a bait protein fused to a DNA
encoding a DNA binding moiety derived from the bacterial
5 LexA protein. The use of a LexA DNA binding domain
provides certain advantages. For example, in yeast, the
LexA moiety contains no activation function and has no
known effect on transcription of yeast genes (Brent and
Ptashne, Nature 312:612-615, 1984; Brent and Ptashne,
10 Cell 43:729-736, 1985). In addition, use of the LexA
rather than, for example, the GAL4 DNA-binding domain
allows conditional expression of prey proteins in
response to galactose induction; this facilitates
detection of prey proteins that might be toxic to the
15 host cell if expressed continuously. Finally, the use of
a well-defined system, such as LexA, allows knowledge
regarding the interaction between LexA and the LexA
binding site (i.e., the LexA operator) to be exploited
for the purpose of optimizing operator occupancy and/or
20 optimizing the geometry of the bound bait protein to
effect maximal gene activation.

      Preferably, the bait protein also includes a LexA
dimerization domain; this optional domain facilitates
efficient LexA dimer formation. Because LexA binds its
25 DNA binding site as a dimer, inclusion of this domain in
the bait protein also optimizes the efficiency of
operator occupancy (Golemis and Brent, Mol. Cell Biol.
12:3006-3014, 1992).

      LexA represents a preferred DNA binding domain in
30 the invention. However, any other transcriptionally-
inert or essentially transcriptionally-inert DNA binding
domain may be used in the interaction trap system; such
DNA binding domains are well known and include the DNA
binding portions of the proteins ACE1 (CUP1), lambda cI,
35 lac repressor, jun, fos, GCN4, or the Tet repressor. The

- 20 -

GAL4 DNA binding domain represents a slightly less
preferred DNA binding moiety for the bait proteins.

Bait proteins may be chosen from any protein of
interest and includes proteins of unknown, known, or
5 suspected diagnostic, therapeutic, or pharmacological
importance. Preferred bait proteins include oncoproteins
(such as myc, particularly the C-terminus of myc, ras,
src, fos, and particularly the oligomeric interaction
domains of fos) or any other proteins involved in cell
10 cycle regulation (such as kinases, phosphatases, the
cytoplasmic portions of membrane-associated receptors).
Particular examples of preferred bait proteins include
cyclin and cyclin dependent kinases (for example, Cdk2)
or receptor-ligand pairs, or neurotransmitter pairs, or
15 pairs of other signalling proteins. In each case, the
protein of interest is fused to a known DNA binding
domain as generally described herein. Examples are
provided below using Cdk2 and Ras baits.

Reporters

20 As shown in Figure 1B, one preferred host strain
according to the invention contains two different
reporter genes, the *LEU2* gene and the *lacZ* gene, each
carrying an upstream binding site for the bait protein.
The reporter genes depicted in Figure 1B each include, as
25 an upstream binding site, one or more LexA operators in
place of their native Upstream Activation Sequences
(UASs). These reporter genes may be integrated into the
chromosome or may be carried on autonomously replicating
plasmids (e.g., yeast 2μ plasmids).

30 A combination of two such reporters is preferred
in the in vivo embodiments of the invention for a number
of reasons. First, the *LexAop-LEU2* construction allows
cells that contain interacting proteins to select
themselves by growth on medium that lacks leucine,
35 facilitating the examination of large numbers of

- 21 -

potential candidate interactor protein-containing cells.
Second, the *LexAop-lacZ* reporter allows LEU⁺ cells to be
quickly screened to confirm an interaction. And, third,
among other technical considerations, the *LexAop-LEU2*
5 reporter provides an extremely sensitive first selection,
while the *LexAop-lacZ* reporter allows discrimination
between proteins of different interaction affinities.

        Although the reporter genes described herein
represent a preferred embodiment of the invention, other
10 equivalent genes whose expression may be detected or
assayed by standard techniques may also be employed in
conjunction with, or instead of, the *LEU2* and *lacZ* genes.
Generally, such reporter genes encode an enzyme that
provides a phenotypic marker, for example, a protein that
15 is necessary for cell growth or a toxic protein leading
to cell death, or encoding a protein detectable by a
color assay or because its expression leads to the
presence or absence of color. Alternatively, the
reporter gene may encode a suppressor tRNA whose
20 expression may be assayed, for example, because it
suppresses a lethal host cell mutation. Particular
examples of other useful genes whose transcription can be
detected include amino acid and nucleic acid biosynthetic
genes (such as yeast *HIS3*, *URA3*, *TRP1*, and *LYS2*) *GAL1*, E.
25 coli *galK* (which complements the yeast *GAL1* gene), and
the reporter genes CAT, GUS, florescent proteins and
derivatives thereof, and any gene encoding a cell surface
antigen for which antibodies are available (e.g., CD4).
Reporter genes may be assayed by either qualitative or
30 quantitative means to distinguish candidate interactors
as agonists or antagonists.

Prey proteins

        In the selection described herein, another DNA
construction is utilized which encodes a series of
35 candidate interacting proteins (i.e., prey proteins);

- 22 -

each is conformationally-constrained, either by being
embedded in a conformation-constraining protein or
because the prey protein's amino and carboxy termini are
linked (e.g., by disulfide bonding). An exemplary prey
5   protein includes an invariant N-terminal moiety carrying,
amino to carboxy terminal, an ATG for protein expression,
an optional nuclear localization sequence, a weak
activation domain (e.g., the B112 or B42 activation
domains of Ma and Ptashne; Cell 51:113, 1987), and an
10 optional epitope tag for rapid immunological detection of
fusion protein synthesis. Library sequences, random or
intentionally designed synthetic DNA sequences, or
sequences encoding conformationally-constrained proteins,
may be inserted downstream of this N-terminal fragment to
15 produce fusion genes encoding prey proteins.

Prey proteins other than those described herein
are also useful in the invention. For example, cDNAs may
be constructed from any mRNA population and inserted into
an equivalent expression vector. Such a library of
20 choice may be constructed *de novo* using commercially
available kits (e.g., from Stratagene, La Jolla, CA) or
using well established preparative procedures (see, e.g.,
Current Protocols in Molecular Biology, New York, John
Wiley & Sons, 1987). Alternatively, a number of cDNA
25 libraries (from a number of different organisms) are
publicly and commercially available; sources of libraries
include, e.g., Clontech (Palo Alto, CA) and Stratagene
(La Jolla, CA). It is also noted that prey proteins need
not be naturally occurring full-length polypeptides. In
30 preferred embodiments, prey proteins are encoded by
synthetic DNA sequences, are the products of randomly
generated open reading frames, are open reading frames
synthesized with an intentional sequence bias, or are
portions thereof. Preferably, such short randomly
35 generated sequences encode peptides between 1 (and

preferably, 6) and 60 amino acids in length. In one
particular example, the prey protein includes only an
interaction domain; such a domain may be useful as a
therapeutic to modulate bait protein activity (i.e., as

5 an antagonist or agonist).

Similarly, any number of activation domains may be
used for that portion of the prey molecule; such
activation domains are preferably weak activation
domains, i.e., weaker than the GAL4 activation region II

10 moiety and preferably no stronger than B112 (as measured,
e.g., by a comparison with GAL4 activation region II or
B112 in parallel β-galactosidase assays using *lacZ*
reporter genes); such a domain may, however, be weaker
than B112. In particular, the extraordinary sensitivity

15 of the *LEU2* selection scheme allows even extremely weak
activation domains to be utilized in the invention.
Examples of other useful weak activation domains include
B17, B42, and the amphipathic helix (AH) domains
described in Ma and Ptashne (Cell 51:113, 1987), Ruden

20 et al. (Nature 350:426-430, 1991), and Giniger and
Ptashne (Nature 330:670, 1987).

The prey proteins, if desired, may include other
optional nuclear localization sequences (e.g., those
derived from the *GAL4* or *MATα2* genes) or other optional

25 epitope tags (e.g., portions of the c-myc protein or the
flag epitope available from Immunex). These sequences
optimize the efficiency of the system, but are not
required for its operation. In particular, the nuclear
localization sequence optimizes the efficiency with which

30 prey molecules reach the nuclear-localized reporter gene
construct(s), thus increasing their effective
concentration and allowing one to detect weaker protein
interactions. The epitope tag merely facilitates a
simple immunoassay for fusion protein expression.

- 24 -

Those skilled in the art will also recognize that the above-described reporter gene, DNA binding domain, and gene activation domain components may be derived from any appropriate eukaryotic or prokaryotic source,
5 including yeast, mammalian cell, and prokaryotic cell genomes or cDNAs as well as artificial sequences. Moreover, although yeast represents a preferred host organism for the interaction trap system (for reasons of ease of propagation, genetic manipulation, and large
10 scale screening), other host organisms such as mammalian cells may also be utilized. If a mammalian system is chosen, a preferred reporter gene is the sensitive and easily assayed CAT gene; useful DNA binding domains and gene activation domains may be chosen from those
15 described above (e.g., the LexA DNA binding domain and the B42 or B112 activation domains).

Conformation-Constraining Proteins

According to one embodiment of the present invention, the DNA sequence encoding the prey protein is
20 embedded in a DNA sequence encoding a conformation-constraining protein (i.e., a protein that decreases the flexibility of the amino and carboxy termini of the prey protein). Methods for directly linking the amino and carboxy termini of a protein (e.g., through disulfide
25 bonding of appropriately positioned cysteine residues) are described above. As an alternative to this approach, conformation-constraining proteins may be utilized. In general, conformation-constraining proteins act as scaffolds or platforms, which limit the number of
30 possible three dimensional configurations the peptide or protein of interest is free to adopt. Preferred examples of conformation-constraining proteins are thioredoxin or other thioredoxin-like sequences, but many other proteins are also useful for this purpose. Preferably,
35 conformation-constraining proteins are small in size

- 25 -

(generally, less than or equal to 200 amino acids), rigid
in structure, of known three dimensional configuration,
and are able to accommodate insertions of proteins of
interest without undue disruption of their structures. A
5 key feature of such proteins is the availability, on
their solvent exposed surfaces, of locations where
peptide insertions can be made (e.g., the thioredoxin
active-site loop). It is also preferable that
conformation-constraining protein producing genes be
10 highly expressible in various prokaryotic and eukaryotic
hosts, or in suitable cell-free systems, and that the
proteins be soluble and resistant to protease
degradation. Examples of conformation-constraining
proteins useful in the invention include nucleases (e.g.,
15 RNase A), proteases (e.g., trypsin), protease inhibitors
(e.g., bovine pancreatic trypsin inhibitor), antibodies
or rigid fragments thereof, and conotoxins. This list,
however, is not limiting. It is expected that other
conformation-constraining proteins having sequences not
20 identified above, or perhaps not yet identified or
published, may be useful based upon their structural
stability and rigidity.

        As mentioned above, one preferred conformation-
constraining protein according to the invention is
25 thioredoxin or other thioredoxin-like proteins. As one
example of a thioredoxin-like protein useful in this
invention, E. coli thioredoxin has the following
characteristics. E. coli thioredoxin is a small protein,
only 11.7 kD, and can be produced to high levels. The
30 small size and capacity for high level synthesis of the
protein contributes to a high intracellular
concentration. E. coli thioredoxin is further
characterized by a very stable, tight tertiary structure
which can facilitate protein purification.

- 26 -

The three dimensional structure of E. coli thioredoxin is known and contains several surface loops, including a distinctive Cys....Cys active-site loop between residues Cys_{33} and Cys_{36} which protrudes from the

5   body of the protein. This Cys....Cys active-site loop is an identifiable, accessible surface loop region and is not involved in interactions with the rest of the protein which contribute to overall structural stability. It is therefore a good candidate as a site for prey protein

10 insertions. Human thioredoxin, glutaredoxin, and other thioredoxin-like molecules also contain this Cys....Cys active-site loop. Both the amino- and carboxyl-termini of E. coli thioredoxin are on the surface of the protein and are also readily accessible for fusion construction.

15 E. coli thioredoxin is also stable to proteases, stable in heat up to 80°C and stable to low pH.

      Other thioredoxin-like proteins encoded by thioredoxin-like DNA sequences useful in this invention share homologous amino acid sequences, and similar

20 physical and structural characteristics. Thus, DNA sequences encoding other thioredoxin-like proteins may be used in place of E. coli thioredoxin according to this invention. For example, the DNA sequence encoding other species' thioredoxin, e.g., human thioredoxin, are

25 suitable. Human thioredoxin has a three-dimensional structure that is virtually superimposable on E. coli's three-dimensional structure, as determined by comparing the NMR structures of the two molecules. Forman-Kay et al., Biochem. 30:2685 (1991). Human thioredoxin also

30 contains an active-site loop structurally and functionally equivalent to the Cys....Cys active-site loop found in the E. coli protein. It can be used in place of or in addition to E. coli thioredoxin in the production of protein and small peptides in accordance

35 with the method of this invention. Insertions into the

- 27 -

human thioredoxin active-site loop and onto the amino
terminus may be as well-tolerated as those in E. coli
thioredoxin.

Other thioredoxin-like sequences which may be
5   employed in this invention include all or portions of the
proteins glutaredoxin and various species' homologs
thereof (Holmgren, supra). Although E. coli glutaredoxin
and E. coli thioredoxin share less than 20% amino acid
homology, the two proteins do have conformational and
10 functional similarities (Eklund et al., EMBO J. 3:1443-
1449 (1984)) and glutaredoxin contains an active-site
loop structurally and functionally equivalent to the
Cys....Cys active-site loop of E. coli thioredoxin.
Glutaredoxin is therefore a thioredoxin-like molecule as
15 defined herein.

In addition, the DNA sequence encoding protein
disulfide isomerase (PDI), or that portion containing the
thioredoxin-like domain, and its various species'
homologs thereof (Edman et al., Nature 317:267-270
20 (1985)) may also be employed as a thioredoxin-like DNA
sequence, since a repeated domain of PDI shares >30%
homology with E. coli thioredoxin and that repeated
domain contains an active-site loop structurally and
functionally equivalent to the Cys....Cys active-site
25 loop of E. coli thioredoxin. The two latter publications
are incorporated herein by reference for the purpose of
providing information on glutaredoxin and PDI which is
known and available to one of skill in the art.

Similarly the DNA sequence encoding
30 phosphoinositide-specific phospholipase C (PI-PLC),
fragments thereof, and various species' homologs thereof
(Bennett et al., Nature, 334:268-270 (1988)) may also be
employed in the present invention as a thioredoxin-like
sequence based on the amino acid sequence homology with
35 E. coli thioredoxin, or alternatively based on similarity

- 28 -

in three dimensional conformation and the presence of an
active-site loop structurally and functionally equivalent
to Cys....Cys active-site loop of E. coli thioredoxin.
All or a portion of the DNA sequence encoding an
5 endoplasmic reticulum protein, ERp72, or various species
homologs thereof are also included as thioredoxin-like
DNA sequences for the purposes of this invention
(Mazzarella et al., J. Biol. Chem. 265:1094-1101 (1990))
based on amino acid sequence homology, or alternatively
10 based on similarity in three dimensional conformation and
the presence of an active-site loop structurally and
functionally equivalent to Cys....Cys active-site loop of
E. coli thioredoxin. Another thioredoxin-like sequence
is a DNA sequence which encodes all or a portion of an
15 adult T-cell leukemia-derived factor (ADF) or other
species homologs thereof (Wakasugi et al., Proc. Natl.
Acad. Sci. USA, 87:8282-8286 (1990)). ADF is now
believed to be human thioredoxin. Similarly, the protein
responsible for promoting disulfide bond formation in the
20 periplasm of E. coli, the product of the dsbA gene
(Bardwell et al., Cell 67:581-89, 1991) also can be
considered a thioredoxin-like sequence. The three latter
publications are incorporated herein by reference for the
purpose of providing information on PI-PLC, ERp72, ADF,
25 and dsbA which are known and available to one of skill in
the art.

       It is expected from the definition of thioredoxin-
like sequences used above that other sequences not
specifically identified above, or perhaps not yet
30 identified or published, may be useful as thioredoxin-
like sequences based on their amino acid sequence
homology to E. coli thioredoxin or based on having three
dimensional structures substantially similar to E. coli
or human thioredoxin and having an active-site loop
35 functionally and structurally equivalent to the

- 29 -

Cys....Cys active-site loop of E. coli thioredoxin. One
skilled in the art can determine whether a molecule has
these latter two characteristics by comparing its three-
dimensional structure, as analyzed for example by x-ray
5 crystallography or two-dimensional NMR spectroscopy, with
the published three-dimensional structure for E. coli
thioredoxin and by analyzing the amino acid sequence of
the molecule to determine whether it contains an active-
site loop that is structurally and functionally
10 equivalent to the Cys....Cys active-site loop of E. coli
thioredoxin. By "substantially similar" in three-
dimensional structure or conformation is meant as similar
to E. coli thioredoxin as is glutaredoxin. In addition a
predictive algorithm has been described which enables the
15 identification of thioredoxin-like proteins via computer-
assisted analysis of primary sequence (Ellis et al.,
Biochemistry 31:4882-91 (1992)). Based on the above
description, one of skill in the art will be able to
select and identify, or, if desired, modify, a
20 thioredoxin-like DNA sequence for use in this invention
without resort to undue experimentation. For example,
simple point mutations made to portions of native
thioredoxin or native thioredoxin-like sequences which do
not effect the structure of the resulting molecule are
25 alternative thioredoxin-like sequences, as are allelic
variants of native thioredoxin or native thioredoxin-like
sequences.

     DNA sequences which hybridize to the sequence for
E. coli thioredoxin or its structural homologs under
30 either stringent or relaxed hybridization conditions also
encode thioredoxin-like proteins for use in this
invention. An example of one such stringent
hybridization condition is hybridization at 4X SSC at
65°C, followed by a washing in 0.1X SSC at 65°C for an
35 hour. Alternatively an exemplary stringent hybridization

condition is in 50% formamide, 4X SSC at 42°C. Examples
of non-stringent hybridization conditions are 4X SSC at
50°C or hybridization with 30-40% formamide at 42°C. The
use of all such thioredoxin-like sequences are believed
5 to be encompassed in this invention.

It may be preferred for a variety of reasons that
prey proteins be fused within the active-site loop of
thioredoxin or thioredoxin-like molecules. The face of
thioredoxin surrounding the active-site loop has evolved,
10 in keeping with the protein's major function as a
nonspecific protein disulfide oxido-reductase, to be able
to interact with a wide variety of protein surfaces. The
active-site loop region is found between segments of
strong secondary structure and this provides a rigid
15 platform to which one may tether prey proteins.

A small prey protein inserted into the active-site
loop of a thioredoxin-like protein is present in a region
of the protein which is not involved in maintaining
tertiary structure. Therefore the structure of such a
20 fusion protein is stable. Indeed, <u>E. coli</u> thioredoxin
can be cleaved into two fragments at a position close to
the active-site loop, and yet the tertiary interactions
stabilizing the protein remain.

The active-site loop of <u>E. coli</u> thioredoxin has
25 the sequence NH_2...Cys_{33}-Gly-Pro-Cys_{36}...COOH. Fusing a
selected prey protein with a thioredoxin-like protein in
the active loop portion of the protein constrains the
prey at both ends, reducing the degrees of conformational
freedom of the prey protein, and consequently reducing
30 the number of alternative structures taken by the prey.
The inserted prey protein is bound at each end by
cysteine residues, which may form a disulfide linkage to
each other as they do in native thioredoxin and further
limit the conformational freedom of the inserted prey.

In addition, by being positioned within the
active-site loop, the prey protein is placed on the
surface of the thioredoxin-like protein, an advantage for
use in screening for bioactive protein conformations and
5   other assays. In general, the utility of thioredoxin or
other thioredoxin-like proteins is described in McCoy et
al., U.S. Pat. No. 5,270,181 and LaVallie et al.,
Bio/Technology 11:187-193 (1993). These two references
are hereby incorporated by reference.

10   There now follows a description of thioredoxin
interaction trap systems according to the invention.
These examples are designed to illustrate, not limit, the
invention.

Thioredoxin Interaction Trap System

15   Interaction trap systems utilizing
conformationally-constrained proteins have been developed
for the detection of protein interactions, for the
identification and isolation of proteins participating in
such interactions, and for the identification and
20   isolation of agonists and antagonists of such
interactions. Exemplary systems are now described.

1.   Thioredoxin Interaction Trap with Cdk2 bait
        Progression of eukaryotic cells through the cell
cycle requires the coordinated action of a number of
25   regulatory proteins that interact with and regulate the
activity of Cdks (Sherr, Cell 79:551-555 (1994)). These
modulatory proteins include cyclins, which positively
regulate Cdk activity, Cyclin Dependent kinase inhibitors
(Ckis), and a number of protein kinases and phosphatases,
30   some of which, such as CAK and Cdc25, positively regulate
kinase activity, some of which, such as Wee1, inhibit
kinase activity, and some of which, such as Cdi1 (Gyuris
et al., Cell 75:791-803 (1993)), have effects that are so

- 32 -

far unknown (reviewed in Morgan, Nature 374:131-134
(1995)). Cdk2 is thought to be required for higher
eukaryotic cells to progress from G1 into S-phase (Fang &
Newport, J. Cell Biol. 66:731-742 (1991); Pagano et al.
5 J. Cell Biol. 121:101-111 (1993); van den Heuvel &
Harlow, Science 262: 2050-2054 (1993)). Cdk2 kinase
activity is positively regulated by Cyclin E and Cyclin A
(Koff et al., Science 257:1689-1694 (1992); Dulic et al.,
Science 257:1958-1961 (1992); Tsai et al., Nature
10 353:174-7 (1991)), negatively regulated by p21, p27 and
p57 (Harper et al., Cell 75:805-816 (1993); Polyak et
al., Genes Dev. 8:9-22 (1994); Toyoshima & Hunter, Cell
78:67-74 (1994); Matsuoka et al. Genes Dev. 9:660-662
(1995); Lee et al., Genes Dev. 9:639-649 (1995)); in
15 addition, Cdk2 complexes with Cdi1 at the G1 to S
transition (Gyuris et al., Cell 75:791-803 (1993)). Here
we describe the use of a yeast two-hybrid system to
select molecules which recognize Cdk2 from combinatorial
libraries.

20   A prey vector is constructed containing the E.
coli thioredoxin gene (trxA). pJG 4-4 (Gyuris et al.,
Cell 75:791, 1993) is used as the vector backbone and cut
with EcoRI and XhoI. A DNA fragment encoding the B112
transcription activation domain is obtained by PCR
25 amplification of plasmid LexA-B112 (Doug Ruden, Ph.D.
thesis, Harvard University, 1992) and cut with MunI and
NdeI. The E. coli trxA gene is excised from the vector
pALTRXA-781 (U.S. Pat. No. 5,292,646; InVitrogen Corp.,
San Diego, CA) by digestion with NdeI and SalI. The trxA
30 and B112 fragments are then ligated by standard
techniques into the EcoRI/XhoI-cut pJG 4-4 backbone,
forming pYENAeTRX. This vector encodes a fusion protein
comprising the SV40 nuclear localization domain, the B112
transcription activation domain, an hemagglutinin epitope
35 tag, and E. coli thioredoxin (Fig. 2A).

- 33 -

Peptide libraries are constructed as follows. The DNA oligomer 5' GACTGACTGGTCCG(NNK)$_{20}$GGTCCTCAGTCAGTCAG 3' (with N = A, C, G, T and K = G, T) (SEQ ID NO: 4) is synthesized and annealed to the second oligomer (5'

5 CTGACTGACTGAGGACC 3') (SEQ ID NO: 5) in order to form double stranded DNA at the 3' end of the first oligomer. The second strand is enzymatically completed using Klenow enzyme, priming synthesis with the second oligomer. The product is cleaved with AvaII, and inserted into RsrII

10 cut pYENAeTRX. After ligation, the construct is used to transform E. coli by standard methods (Ausubel et al., supra). The library contained 2.9 x 109 members, of which more than 109 directs the synthesis of peptides.

To screen for interacting peptides, 20 μg of the

15 library is used to transform the yeast strain EGY48 (*Matα his3 leu2::2Lexop-LEU2 ura3 trp1 LYS2*; Gyuris et al., supra). This strain also contains the reporter plasmid pSH 18-34, a pLR1Δ1 derivative, containing the yeast 2μ replication origin, the *URA3* gene, and a *GAL1-lacZ*

20 reporter gene with the *GAL1* upstream regulatory elements replaced with 4 colE1 LexA operators (West et al., Mol. Cell Biol. 4:2467, 1984; Ebina et al., J. Biol. Chem. 258:13258, 1983; Hanes and Brent, Cell 57:1275, 1989), as well as the bait vector pLexA202-Cdk2 (Cdk2 encodes the

25 human cyclin dependent kinase 2, an essential cell cycle enzyme) (Gyuris et al., supra; Tsai et al., Oncogene 8:1593, 1993). About 2.5 x 10^6 transformants are obtained and pooled. The first selection step, growth on leucine-deficient medium after induction with 2%

30 galactose/1% raffinose (Gyuris et al., supra; Guthrie and Fink, Guide to Yeast Genetics and Molecular Biology, Vol. 194, 1991), is performed with an 8-fold redundancy (20 x 10^6 cfu) of the library in yeast, and about 900 colonies are obtained after growth at 30°C for 5 days. The 300

35 largest colonies are streak purified and tested for the

galactose-dependent expression of the *LEU2* gene product
and of β-galactosidase (encoded by pSH 18-34), the latter
giving rise to blue yeast colonies in the presence of
Xgal in the medium (Ausubel et al., supra). Thirty-three
5 colonies fulfill these requirements which, after
sequencing, include 14 different clones all of which bind
specifically to a LexA-Cdk2 bait, but not to LexA or to a
LexA-Cdk3 bait (Finley et al., Proc. Natl. Acad. Sci.,
1994). The strength of binding is judged according to
10 the intensity of the blue color formed by a colony of the
yeast that contains each different interactor. By this
means, each interactor is classified as a strong, medium,
or weak binder, which is normalized to the amount of blue
color caused by the various naturally-occurring partner
15 proteins of Cdk2 in side by side mating interaction
assays. An example of the peptide sequence of one
representative of each class is given here:

Strong binder: peptide 3 (SEQ ID NO: 6)
-Gly$_{34}$-Pro$_{35}$-Leu-Val-Cys-Lys-Ser-Tyr-Arg-Leu-Asp-Trp-
20 Glu-Ala-Gly-Ala-Leu-Phe-Arg-Ser-Leu-Phe-Gly$_{34}$-Pro$_{35}$-

Medium binder: peptide 2 (SEQ ID NO: 7)
-Gly$_{34}$-Pro$_{35}$-Met-Val-Val-Ala-Ala-Glu-Ala-Val-Arg-Thr-
Val-Leu-Leu-Ala-Asp-Gly-Gly-Asp-Val-Thr-Gly$_{34}$-Pro$_{35}$-

Weak binder: peptide 6 (SEQ ID NO: 8)
25 -Gly$_{34}$-Pro$_{35}$-Pro-Asn-Trp-Pro-His-Gln-Leu-Arg-Val-Gly-
Arg-Val-Leu-Trp-Glu-Arg-Leu-Ser-Phe-Glu-Gly$_{34}$-Pro$_{35}$-

Control peptides which do not bind detectably are: c4:
Arg-Arg-Ala-Ser-Val-Cys-Gly-Pro-Leu-Leu-Ser-Lys-Arg-Gly-
Tyr-Gly Pro-Pro-Phe-Tyr-Leu-Ala-Gly-Met-Thr-Ala-Pro-Glu-
30 Gly-Pro-Cys (SEQ ID NO: 14) and c: Arg-Arg-Ala-Ser-Val-
Cys-Gly-Pro-Leu-His-Tyr-Trp-Gly-Leu-Gly-Gly-Phe-Val-Asp-

Leu-Trp-Gln-Glu-Thr-Thr-Gly-Val-Gly-Pro-Cys (SEQ ID NO: 15).

        Figure 3A shows that 5 of the peptides reacted strongly with the LexA-Cdk2 bait but not with a large
5 number of unrelated proteins. None of the Cdk2 aptamers interacted with CDC28 or Cdc2, which are both 65% identical to Cdk2. However, 2 of the 5 Cdk2 interactors also interacted with human Cdk3, and 1 of the 5 also interacted with Drosophila Cdc2c, suggesting that these
10 peptides recognize determinants common to these proteins. Both theoretical considerations and calibration experiments with lambda repressor's C terminus suggest that transcription of the pSH18-34 reporter in EGY48 can be activated by protein interactions with Kds as weak as
15 10^{-6}M. The fact that peptides 3 and 13 direct robust transcription of the this *LexAop-lacZ* reporter is consistent with the idea that they may interact significantly more tightly. The sequence of these peptides is shown in Figure 3B. Two of the peptides are
20 longer than unit length; both are apparently artifacts of the *in vitro* manipulations used to construct the library. No peptide showed significant sequence similarity to known proteins, and none showed more than random similarity to any other, suggesting that we have not
25 exhausted the peptide motifs capable of recognizing Cdk2.
        To confirm the specificity of the Cdk2 interaction, we immobilized a Gst-Cdk2 fusion protein on glutathione sepharose beads, and used these beads to specifically precipitate two bacterially expressed
30 peptide aptamers (Fig. 4). Gst-Cdk2 was expressed in *E. coli* and purified on glutathione sepharose as described. (Lee et al., Nature 374: 91-94 (1995)). Peptides 3 and 13 were made as follows: fragments that directed the synthesis of peptides 3 and 13 were made by PCR
35 amplification of the insert encoded by the corresponding

library plasmid and introduced into pAL-TrxA (LaVallie et
al., Bio/Technology 11:187-193 (1993)). Fusion proteins
were expressed and lysed in a French pressure cell as
described (LaVallie et al., *BIO/Technology* 11, 187-193
5   (1993)). Coprecipitation with Gst-Sepharose beads was
done as described (Lee et al., *Nature* 374, 91-94 (1995)),
and samples were run on a 15% SDS polyacrylamide gel and
transferred to nylon membranes. TrxA-containing fusion
proteins were visualized by probing the membranes with an
10 anti-TrxA antibody, and developing the immobilized
antibody with peroxidase-coupled anti-rabbit IgG antibody
ECL reagents according to the manufacturer's instructions
(Amersham, Arlington Heights, IL).
        These experiments demonstrate that the
15 interactions between Cdk2 and the peptide aptamers can be
observed *in vitro*, and is thus independent of any bridge
proteins native to yeast. Once identified, these
peptides may be used in competition experiments.
        The ability to select TrxA-peptides that interact
20 specifically with designated intracellular baits allows
for the creation of other classes of intracellular
reagents. For example, appropriately derivitized TrxA-
peptide fusions may allow the creation of antagonists or
agonists (as described above). Alternatively, peptide
25 fusions allow for the creation of homodimeric or
heterodimeric "matchmakers," which force the interaction
of particular protein pairs. In one particular example,
two proteins are forced together by utilizing a leucine
zipper sequence attached to a conformation-constraining
30 protein containing a candidate interaction peptide. This
protein can bind to both members of a protein pair of
interest and direct their interaction. Alternatively,
the "matchmaker" may include two different sequences, one
having affinity for a first polypeptide and the second
35 having affinity for the second polypeptide; again, the

- 37 -

result is directed interaction between the first and
second polypeptides. Another practical application for
the peptide fusions described herein is the creation of
"destroyers," which target a bound protein for

5 destruction by host proteases. In an example of the
destroyer application, a protease is fused to one
component of an interacting pair and that component is
allowed to interact with the target to be destroyed
(e.g., a protease substrate). By this method, the

10 protease is delivered to its desired site of action and
its proteolytic potential effectively enhanced. Yet
another application of the fusion proteins described
herein are as "conformational stabilizers," which induce
target proteins to favor a particular conformation or

15 stabilize that conformation. In one particular example,
the ras protein has one conformation that signals a cell
to divide and another conformation that signals a cell
not to divide. By selecting a peptide or protein that
stabilizes the desired conformation, one can influence

20 whether a cell will divide. Other proteins that undergo
conformational changes which increase or decrease
activity can also be bound to an appropriate
"conformational stabilizer" to influence the property of
the desired protein.

25 2. Functional Inhibition of Cdk2
                To determine whether Cdk2 interacting peptides
might inhibit Cdk2 function *in vivo*, we took advantage of
the fact that human Cdk2 can complement temperature
sensitive alleles of Cdc28 (Elledge and Spottswood, EMBO

30 10:2653-2659, 1991; Ninomiya et al., PNAS 88:9006-9010,
1991; Meyerson et al., EMBO 11:2909-2917, 1992). Peptide
13 inhibits the plating efficiency of a Cdk2-dependent
yeast. A strain carrying the temperature sensitive
cdc28-1N mutation can form colonies at high temperature

- 38 -

if it carries a plasmid that expresses Cdk2. At the
restrictive temperature, compared to the plating
efficiency of yeast expressing control peptides,
expression of peptide 13 diminishes the plating
5   efficiency of this strain by 10-fold. Both peptide 3 and
13 have similar effects on the plating efficiency at 37°C
of a Cdk2(+) strain that carries the *cdc28-13ts* allele.

Expression of peptide 13 slows the doubling time
of a Cdk2(+), *cdc28ts-1N* strain by a factor of 50%.
10 Microscopic examination of strains expressing the peptide
revealed that a high proportion of these cells had an
elongated morphology characteristic of *cdc28-1N* cells at
the restrictive temperature, whereas cells expressing a
control peptide had a more normal morphology.
15   Peptide 13 does not affect the growth of a cdc28-
1Nts strain at high temperature when the defect is
complemented by a plasmid expressing wild-type Cdc28
product, and has no effect on yeast at the permissive
temperature. While we do not intend to be bound by any
20 particular theory, it appears that this peptide blocks
yeast cell cycle progression by binding to some face of
the Cdk2 molecule and inhibiting its function and thereby
interfering with its ability to interact with cyclins,
other partners, or with substrates.
25 3. Thioredoxin Interaction Trap with OncoRas Bait

The *ras* proteins are essential for many signal
transduction pathways and regulate numerous physiological
functions including cell proliferation. The *ras* genes
were first identified from the genome of Harvey and
30 Kirsten sarcoma virus. The three types of mammalian *ras*
genes (N-, K-*ras*, and H-*ras*) encode highly conserved
membrane-bound guanine nucleotide binding proteins with a
molecular mass of 21 kDa, which cycle between the active
(GTP-bound) form and the inactive (GDP-bound) form.

- 39 -

In normal cells, the active form of Ras is short-
lived, as its intrinsic GTPase activity rapidly converts
the bound-GTP to GDP. The GTPase activity is stimulated
10^5-fold by GTPase-activating proteins (GAPs). GTP-bound
5 Ras interact with GAP, c-Raf, neurofibromatosis type 1
(NF-1) and Ral guanine nucleotide dissociation stimulator
(RalGDS).

Mutationally-activated RAS proteins are found in
about 30% of human tumor cells and have greatly decreased
10 GTPase activity which can not be stimulated by GAPs. The
majority of mutations studied thus far are due to a point
mutation at either residue Gly-12 or residue Gln-61 of
Ras. These Ras mutants remain in the active form and
interact with the downstream effectors to result in
15 tumorigenesis. It has been shown that there are
significant conformational differences between GTP-bound
forms of wild-type and oncogenic RAS proteins. Such
conformational differences are likely causes for
malignant transformation induced by oncogenic ras
20 proteins.

Such mutationally-activated conformational changes
in GTP-bound H-ras mutants provide targets for members of
a conformationally constrained random peptide library.
In the present example, the library is a conformationally-
25 constrained thioredoxin peptide library, as described
above. Library members, which interact with oncogenic
Ras have been identified using a variation of the
interaction trap technology provided above. The
oncogenic Ras peptide aptamers isolated may be assayed
30 for their ability to disrupt the interaction of oncogenic
Ras with known effectors and to inhibit cellular
transformation.

We have used well-characterized oncogenic H-
ras(G12V) for isolation and characterization of its
35 peptide aptamers. Peptide aptamers for other oncogenes

- 40 -

can be isolated using adaptations of this protocol as
provided herein.

Bait Construction

      Construction of LexA-Ras(G12V)/pEG202:H-Ras(G12V)
5 DNA was performed by digesting BTM116-H-Ras(G12V) (Fig.
5) with BamHI and SalI. H-Ras(G12V) DNA was ligated with
pEG202 backbone digested with BamHI and SalI. The
resulting plasmid was called pEG202-H-Ras(G12V) (or V6)
(Fig. 6).

10 *Screening for H-Ras(G12V) peptide aptamers*

      pEG202-H-Ras(G12V) (V6) was transformed into the
EGY48 strain according to a standard yeast transformation
protocol; in particular, the protocol provided by Zymo
Research (Orange County, CA) was used here. EGY48 was
15 grown in YPD medium to OD_{600}=0.2-0.7. Cells were pelleted
at 500 X g for 4 min. and resuspended in 10 ml of EZ1
solution (Zymo Research). The cells were then pelleted
by centrifugation and resuspended in 1 ml of EZ2 (Zymo
Research). Aliquots of competent cells (50 μl) were
20 stored in a -70°C freezer.

      An aliquot of competent cells was mixed with 0.1
μg of LexA-H-Ras(G12V)/pEG202 and 500 μl of EZ3 solution
(Zymo Research). The mixture was incubated at 30°C for
30 min. and plated onto a yeast medium lacking histidine
25 and uracil. One colony was picked and inoculated into
100 ml of glucose Ura⁻His⁻ medium at 30°C with shaking
(150 rpm) until the OD_{600} measurement was 0.96. The
culture was centrifuged at 2000g for 5 min and cell
pellets were resuspended in 5 ml of sterile LiOAc/TE.
30 The cells were again centrifuged as above and resuspended
in 0.5 ml of sterile LiOAc/TE.

      Aliquots (50 μl) of the cells were then incubated
at 30°C for 30 min. with 1 μg of thioredoxin peptide
library DNA, 70 μg of salmon sperm DNA, and 300 μl of
35 sterile 40% PEG 4000 in LiOAc/TE. The mixtures were

heat-shocked at 42°C for 15 min. Each aliquot was plated
onto a 24 cm x 24 cm plate containing glucose Ura⁻His⁻Trp⁻
medium and was incubated at 30°C for two days. The
transforming efficiency typically ranged from 50,000 to
5   100,000 colony forming units per µg of library DNA.

A total of 1.5 million transformants were obtained
and were plated onto the selection medium of
galactose/raffinose Leu⁻Ura⁻His⁻Trp⁻. Of the 338 colonies
formed, among them 50 were randomly picked and inoculated
10   into 5 ml of glucose Leu⁻Ura⁻His⁻Trp⁻ medium for
preparation of yeast plasmid DNA. A half ml of each
yeast culture was mixed with an equal volume of acid-
washed sand and phenol/chloroform/isoamyl alcohol
(24:24:1), and vortexed in a vortexer for 2 min. The
15   mixture was then centrifuged for 15 min., and the
supernatant was precipitated with ethanol. DNA pellets
were resuspended in 50 µl of TE.

One µl of each sample was used to transform *E.
coli* KC8 cells by electroporation. Bacterial
20   transformants were selected on minimal agar supplemented
with uracil, leucine, histidine, and ampicillin. Each
type transformant resulted in final isolation of plasmid
which a leucine marker, which carries a DNA fragment
encoding thioredoxin-peptide fusion protein.

25   Sequence determination of the 50 isolates was
carried out according to the directions of the fmolDNA™
sequencing systems (Promega, Madison, WI) using primer
5'-GACGGGGCGATCCTCGTCG-3' (SEQ ID NO:16). Nine out of 50
isolates (referred to as #4, #18, #39, #41, #22, #24,
30   #30, #31, #46) contained unique peptide encoding
sequences, as determined by electrophoresis of the dT/ddT
termination reaction. Among them, the predicted peptide
aptamer sequence of #39 is as follows:
Trp-Ala-Glu-Trp-Cys-Gly-Pro-Val-Cys-Ala-His-Gly-Ser-Arg-
35   Ser-Leu-Thr-Leu-Leu-Thr-Lys-Tyr-His-Val-Ser-Phe-Leu-Gly-

- 42 -

Pro-Cys-Lys-Met-Ile-Ala-Pro-Ile-Leu-Asp (SEQ ID NO:17).
From our results, it appears that approximately 60 unique
H-Ras(G12V) peptide aptamers (338 x 9/50) were isolated
in the first round of screening.

5   Other Embodiments

         As described above, the invention features a
method for detecting and analyzing protein-protein
interactions. Typically, in the above experiments, the
bait protein is fused to the DNA binding domain, and the
10 prey protein (in association with the conformation-
constraining protein) is fused to the gene activation
domain. The invention, however, is readily adapted to
other formats. For example, the invention also includes
a "reverse" interaction trap in which the bait protein is
15 fused to a gene activation domain, and the prey protein
(in association with a conformation-constraining protein)
is fused to the DNA binding domain. Again, an
interaction between the bait and prey proteins results in
activation of reporter gene expression. Such a "reverse"
20 interaction trap system, however, depends upon the use of
prey proteins which do not themselves activate downstream
gene expression.

         The protein interaction assays described herein
can also be accomplished in a cell-free, in vitro system.
25 Such a system begins with a DNA construct including a
reporter gene operably linked to a DNA-binding-protein
recognition site (e.g., a LexA binding site). To this
DNA is added a bait protein (e.g., any of the bait
proteins described herein bound to a LexA DNA binding
30 domain) and a prey protein (e.g., one of a library of
conformationally-constrained candidate interactor prey
proteins bound to a gene activation domain). Interaction
between the bait and prey protein is assayed by measuring
the reporter gene product, either as an RNA product, as

- 43 -

an in vitro translated protein product, or by some
enzymatic activity of the translated reporter gene
product. This in vitro system may also be used to
identify agonists or antagonists, simply by adding to a

5 known pair of interacting proteins (in the above
described system) a candidate agonist or antagonist
interactor and assaying for an increase or decrease
(respectively) in reporter gene expression, as compared
to a control reaction lacking the candidate compound or

10 protein. To facilitate large scale screening, candidate
prey proteins or candidate agonists or antagonists may be
initially tested in pools, for example, of ten or twenty
candidate compounds or proteins. From pools
demonstrating a positive result, the particular

15 interacting protein or agonist or antagonist is then
identified by individually assaying the components of the
pool. Such in vitro systems are amenable to robotic
automation or to the production of kits. Kits including
the components of any of the interaction trap systems

20 described herein are also included in the invention.

The components (e.g., the various fusion proteins
or DNA therefor) of any of the in vivo or in vitro
systems of the invention may be provided sequentially or
simultaneously depending on the desired experimental

25 design.

- 44 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT: Brent, Roger
                    McCoy, John M.
                    Jessen, Timm H.
                    Xu, Chanxing Wilson

    (ii) TITLE OF INVENTION: INTERACTION TRAP SYSTEMS FOR
                                        DETECTING PROTEIN INTERACTIONS

    (iii) NUMBER OF SEQUENCES: 20

    (iv) CORRESPONDENCE ADDRESS:

        (A) ADDRESSEE: Fish & Richardson, P.C.
        (B) STREET: 225 Franklin Street
        (C) CITY: Boston
        (D) STATE: Massachusetts
        (E) COUNTRY: USA
        (F) ZIP: 02110-2804

    (v) COMPUTER READABLE FORM:

        (A) MEDIUM TYPE: Floppy disk
        (B) COMPUTER: IBM PC compatible
        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
        (D) SOFTWARE: PatentIn Release #1.0, Version
                            #1.30

    (vi) CURRENT APPLICATION DATA:

        (A) APPLICATION NUMBER:
        (B) FILING DATE:
        (C) CLASSIFICATION:

    (vii) PRIOR APPLICATION DATA:

        (A) APPLICATION NUMBER:
        (B) FILING DATE:

    (viii) ATTORNEY/AGENT INFORMATION:

        (A) NAME: Paul T. Clark
        (B) REGISTRATION NUMBER: 30,162
        (C) REFERENCE/DOCKET NUMBER: 00786/288001

    (ix) TELECOMMUNICATION INFORMATION:

        (A) TELEPHONE: (617) 542-5070
        (B) TELEFAX: (617) 542-8906
        (C) TELEX: 200154

(2) INFORMATION FOR SEQ ID NO:1:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 20
        (B) TYPE: amino acid
        (C) STRANDEDNESS:

- 45 -

        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Leu Val Cys Lys Ser Tyr Arg Leu Asp Trp Glu Ala Gly Ala Leu Phe
1               5                   10                  15

Arg Ser Leu Phe
            20
```

  (2) INFORMATION FOR SEQUENCE ID NO:2:

      (i) SEQUENCE CHARACTERISTICS:

          (A) LENGTH: 20
          (B) TYPE: amino acid
          (C) STRANDEDNESS:
          (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Val Val Ala Ala Glu Ala Val Arg Thr Val Leu Leu Ala Asp Gly
1               5                   10                  15

Gly Asp Val Thr
            20
```

  (2) INFORMATION FOR SEQ ID NO:3:

      (i) SEQUENCE CHARACTERISTICS:

          (A) LENGTH: 20
          (B) TYPE: amino acid
          (C) STRANDEDNESS:
          (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Pro Asn Trp Pro His Gln Leu Arg Val Gly Arg Val Leu Trp Glu Arg
1               5                   10                  15

Leu Ser Phe Glu
            20
```

  (2) INFORMATION FOR SEQ ID NO:4:

      (i) SEQUENCE CHARACTERISTICS:

          (A) LENGTH: 91
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

    (ix) FEATURE:

    (D) OTHER INFORMATION: N is A or T or G or C; K is G or T.

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
GACTGACTGG TCCGNNKNNK NNKNNKNNKN NKNNKNNKNN KNNKNNKNNK NNKNNKNNKN   60
```

- 46 -

NKNNKNNKNN KNNKGGTCCT CAGTCAGTCA G   31

(2) INFORMATION FOR SEQ ID NO:5:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 17
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTGACTGACT GAGGACC   17

(2) INFORMATION FOR SEQ ID NO:6:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 24
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Gly Pro Leu Val Cys Lys Ser Tyr Arg Leu Asp Trp Glu Ala Gly Ala
1   5   10   15

Leu Phe Arg Ser Leu Phe Gly Pro
            20

(2) INFORMATION FOR SEQ ID NO:7:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 24
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Gly Pro Met Val Val Ala Ala Glu Ala Val Arg Thr Val Leu Leu Ala
1   5   10   15

Asp Gly Gly Asp Val Thr Gly Pro
            20

(2) INFORMATION FOR SEQ ID NO:8:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 24
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Pro Pro Asn Trp Pro His Gln Leu Arg Val Gly Arg Val Leu Trp
1   5   10   15

Glu Arg Leu Ser Phe Glu Gly Pro
                20

(2) INFORMATION FOR SEQ ID NO:9:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 20
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Ser Val Arg Met Arg Tyr Gly Ile Asp Ala Phe Phe Asp Leu Gly Gly Leu
1   5   10   15

Leu His Gly
        20

(2) INFORMATION FOR SEQ ID NO:10:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 42
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Glu Leu Arg His Arg Leu Gly Arg Ala Leu Ser Glu Asp Met Val Arg Gly
1   5   10   15

Leu Ala Trp Gly Pro Thr Ser His Cys Ala Thr Val Pro Gly Thr Ser Asp
        20   25   30

Leu Trp Arg Val Ile Arg Phe Leu
35   40

(2) INFORMATION FOR SEQ ID NO:11:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 20
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Tyr Ser Phe Val His His Gly Phe Phe Asn Phe Arg Val Ser Trp Arg Glu
1   5   10   15

- 48 -

Met Leu Ala
        20

(2) INFORMATION FOR SEQ ID NO:12:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 20
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Gln Val Trp Ser Leu Trp Ala Leu Gly Trp Arg Trp Leu Arg Arg Tyr Gly
1   5   10   15

Trp Asn Met
        20

(2) INFORMATION FOR SEQ ID NO:13:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 20
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Trp Arg Arg Met Glu Leu Asp Ala Glu Ile Arg Trp Val Lys Pro Ile Ser
1   5   10   15

Pro Leu Glu
        20

(2) INFORMATION FOR SEQ ID NO:14:

        (i) SEQUENCE CHARACTERISTICS:

                (A) LENGTH: 31
                (B) TYPE: amino acid
                (C) STRANDEDNESS:
                (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Arg Arg Ala Ser Val Cys Gly Pro Leu Leu Ser Lys Arg Gly Tyr Gly
1   5   10   15

Pro Pro Phe Tyr Leu Ala Gly Met Thr Ala Pro Glu Gly Pro Cys
            20   25   30

- 49 -

(2) INFORMATION FOR SEQ ID NO:15:

      (i) SEQUENCE CHARACTERISTICS:

          (A) LENGTH: 30
          (B) TYPE: amino acid
          (C) STRANDEDNESS:
          (D) TOPOLOGY: linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Arg Arg Ala Ser Val Cys Gly Pro Leu His Tyr Trp Gly Leu Gly Gly
1   5   10   15

Phe Val Asp Leu Trp Gln Glu Thr Thr Gly Val Gly Pro Cys
          20   25   30

(2) INFORMATION FOR SEQ ID NO:16:

      (i) SEQUENCE CHARACTERISTICS:

          (A) LENGTH: 19
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GACGGGGCGA TCCTCGTCG   19

(2) INFORMATION FOR SEQ ID NO:17:

      (i) SEQUENCE CHARACTERISTICS:

          (A) LENGTH: 38
          (B) TYPE: amino acid
          (C) STRANDEDNESS:
          (D) TOPOLOGY: linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Trp Ala Glu Trp Cys Gly Pro Val Cys Ala His Gly Ser Arg Ser Leu
1   5   10   15

Thr Leu Leu Thr Lys Tyr His Val Ser Phe Leu Gly Pro Cys Lys Met
          20   25   30

Ile Ala Pro Ile Leu Asp
          35

(2) INFORMATION FOR SEQ ID NO:18:

      (i) SEQUENCE CHARACTERISTICS:

- 50 -

```
              (A) LENGTH:           20
              (B) TYPE:             amino acid
              (C) STRANDEDNESS:
              (D) TOPOLOGY:         linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:


Leu Val Cys Lys Ser Tyr Arg Leu Asp Trp Glu Ala Gly Ala Leu Phe Arg
1               5                   10                  15

Ser Leu Phe
        20
```

```
(2)  INFORMATION FOR SEQ ID NO:19:

        (i) SEQUENCE CHARACTERISTICS:

              (A) LENGTH:           20
              (B) TYPE:             amino acid
              (C) STRANDEDNESS:
              (D) TOPOLOGY:         linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:


Tyr Arg Trp Gln Gln Gly Val Val Pro Ser Asn Trp Ala Ser Cys Ser Phe
1               5                   10                  15

Arg Cys Gly
        20
```

```
(2)  INFORMATION FOR SEQ ID NO:20:

        (i) SEQUENCE CHARACTERISTICS:

              (A) LENGTH:           38
              (B) TYPE:             amino acid
              (C) STRANDEDNESS:
              (D) TOPOLOGY:         linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:


Ser Ser Phe Ser Leu Trp Leu Leu Met Val Lys Ser Ile Lys Arg Ala Ala
1               5                   10                  15

Trp Glu Leu Gly Pro Ser Ser Ala Trp Asn Thr Ser Gly Trp Ala Ser Leu
        20                  25                  30

Ala Asp Phe Tyr
35
```

- 51 -

What is claimed is:

Claims

1.   A method of determining whether a first protein is capable of physically interacting with a second protein, comprising:

   (a) providing a host cell which contains

   (i) a reporter gene operably linked to a DNA-binding-protein recognition site;

   (ii) a first fusion gene which expresses a first fusion protein, said first fusion protein comprising said first protein covalently bonded to a binding moiety which is capable of specifically binding to said DNA-binding-protein recognition site; and

   (iii) a second fusion gene which expresses a second fusion protein, said second fusion protein comprising said second protein covalently bonded to a gene activating moiety and being conformationally-constrained; and

   (b) measuring expression of said reporter gene as a measure of an interaction between said first and said second proteins.

2.   The method of claim 1, wherein said second protein is a short peptide of at least 6 amino acids.

3.   The method of claim 1, wherein said second protein is a short peptide is less than or equal to 60 amino acids in length.

4.   The method of claim 1, wherein said second protein comprises a randomly generated or intentionally designed peptide sequence.

- 52 -

5.   The method of claim 1, wherein said second
protein is conformationally-constrained because it is
covalently bonded to a conformation-constraining protein.

6.   The method of claim 1, wherein said first
5   protein is Cdk2.

7.   The method of claim 1, wherein said first
protein is Ras or an activated Ras.

8.   The method of claim 5, wherein said second
protein is embedded within said conformation-constraining
10   protein.

9.   The method of claim 5, wherein said
conformation-constraining protein is thioredoxin.

10.   The method of claim 5, wherein said
conformation-constraining protein is a thioredoxin-like
15   molecule.

11.   The method of claim 9, wherein said second
protein is inserted into the active site loop of said
thioredoxin protein.

12.   The method of claim 1, wherein said second
20   protein is conformationally-constrained by disulfide
bonds between cysteine residues in the amino-terminus and
in the carboxy-terminus of said second protein.

13.   The method of claim 1, wherein said host cell
is yeast.

25   14.   The method of claim 1, wherein said DNA
binding domain is LexA.

15.   The method of claim 1, wherein said reporter gene is assayed by a color reaction.

16.   The method of claim 1, wherein said reporter gene is assayed by cell viability.

5   17.   A method of detecting an interacting protein in a population of proteins, comprising:
            (a) providing a host cell which contains
            (i) a reporter gene operably linked to a DNA-binding-protein recognition site; and
10   (ii) a fusion gene which expresses a fusion protein, said fusion protein comprising a test protein covalently bonded to a binding moiety which is capable of specifically binding to said DNA-binding-protein recognition site;
15   (b) introducing into said host cell a second fusion gene which expresses a second fusion protein, said second fusion protein comprising one of said population of proteins covalently bonded to a gene activating moiety and being conformationally-constrained; and
20   (c) measuring expression of said reporter gene.

18.   The method of claim 17, wherein said population of proteins comprises short peptides of between 1 and 60 amino acids in length.

25   19.   The method of claim 17, wherein said population of proteins is a set of randomly generated or intentionally designed peptide sequences.

- 54 -

20.   The method of claim 17, wherein said
population of proteins is conformationally-constrained by
covalently bonding to a conformation-constraining
protein.

5   21.   The method of claim 20, wherein each of said
population of proteins is embedded within a conformation-
constraining protein.

22.   The method of claim 20, wherein said
conformation-constraining protein is thioredoxin.

10   23.   The method of claim 22, wherein each of said
population of proteins is inserted into the active site
loop of said thioredoxin.

24.   The method of claim 17, wherein each of said
population of proteins is conformationally-constrained by
15   disulfide bonds between cysteine residues in the amino-
terminus and in the carboxy-terminus of said protein.

25.   The method of claim 17, wherein said first
protein is Cdk2.

26.   The method of claim 17, wherein said first
20   protein is Ras or an activated Ras.

27.   The method of claim 17, wherein said host
cell is yeast.

28.   The method of claim 17, wherein said DNA
binding domain is LexA.

25   29.   The method of claim 17, wherein said reporter
gene is assayed by a color reaction.

- 55 -

30.   The method of claim 17, wherein said reporter gene is assayed by cell viability.

31.   A method of identifying a candidate interactor, comprising:

     (a) providing a reporter gene operably linked to a DNA-binding-protein recognition site;

     (b) providing a first fusion protein, said first fusion protein comprising a first protein covalently bonded to a binding moiety which is capable of specifically binding to said DNA-binding-protein recognition site;

     (c) providing a second fusion protein, said second fusion protein comprising a second protein covalently bonded to a gene activating moiety and being conformationally-constrained, said second protein being capable of interacting with said first protein;

     (d) contacting said candidate interactor with said first protein and/or said second protein; and

     (e) measuring expression of said reporter gene.

32.   The method of claim 31, wherein providing said first fusion protein comprises providing a first fusion gene which expresses said first fusion protein and wherein providing said second fusion protein comprises providing a second fusion gene which expresses said second fusion protein.

33.   The method of claim 31, wherein said first fusion protein and said second fusion protein are permitted to interact prior to contact with said candidate interactor.

- 56 -

34. The method of claim 31, wherein said first fusion protein and said candidate interactor are permitted to interact prior to contact with said second fusion protein.

5   35. The method of claim 31, wherein said candidate interactor is conformationally-constrained.

36. The method of claim 31, wherein said candidate interactor is an antagonist and reduces reporter gene expression.

10   37. The method of claim 31, wherein said candidate interactor is an agonist and increases reporter gene expression.

38. The method of claim 31, wherein said candidate interactor is a member selected from the group
15 consisting of proteins, polynucleotides, and small molecules.

39. The method of claim 31, wherein said candidate interactor is encoded by a member of a cDNA or synthetic DNA library.

20   40. The method of claim 31, wherein said candidate interactor is a mutated form of said first fusion protein or said second fusion protein.

41. The method of claim 32, wherein said reporter gene, said first fusion gene, and said second fusion gene
25 are included on a single piece of DNA.

42. The method of claim 31, wherein said first protein is Cdk2.

43.   The method of claim 31, wherein said first protein is Ras or an activated Ras.

44.   A population of eukaryotic cells, each cell having a recombinant DNA molecule encoding a
5   conformationally-constrained intracellular peptide, there being at least 100 different recombinant molecules in said population, each molecule being in at least one cell of said population.

45.   The population of eukaryotic cells of claim
10   44, wherein said intracellular peptide is conformationally-constrained because it is covalently bonded to a conformation-constraining protein.

46.   The population of claim 45, wherein said intracellular peptide is embedded within said
15   conformation-constraining protein.

47.   The population of eukaryotic cells of claim 45, wherein said conformation-constraining protein is thioredoxin.

48.   The population of eukaryotic cells of claim
20   44, wherein said intracellular peptide is conformationally-constrained by disulfide bonds between cysteine residues in the amino-terminus and in the carboxy-terminus of said second protein.

49.   The population of eukaryotic cells of claim
25   44, wherein said cells are yeast cells.

- 58 -

50. The population of eukaryotic cells of claim 44, wherein said recombinant DNA molecule further encodes a gene activating moiety covalently bonded to said intracellular peptide.

5   51. The population of eukaryotic cells of claim 44, wherein said intracellular peptide physically interacts with a second recombinant protein inside said eukaryotic cells.

52. A method of assaying an interaction between a
10   first protein and a second protein, comprising:
       (a) providing a reporter gene operably linked to a DNA-binding-protein recognition site;
       (b) providing a first fusion protein comprising said first protein covalently bonded to a binding moiety
15   which is capable of specifically binding to said DNA-binding-protein recognition site;
       (c) providing a second fusion protein comprising said second protein covalently bonded to a gene activating moiety and being conformationally-constrained;
20   (d) combining said reporter gene, said first fusion protein, and said second fusion protein; and
       (e) measuring expression of said reporter gene.

53. The method of claim 52, wherein providing said first fusion protein comprises providing a first
25   fusion gene which expresses said first fusion protein and wherein providing said second fusion protein comprises providing a second fusion gene which expresses said second fusion protein.

54. A protein comprising the sequence Leu-Val-
30   Cys-Lys-Ser-Tyr-Arg-Leu-Asp-Trp-Glu-Ala-Gly-Ala-Leu-Phe-Arg-Ser-Leu-Phe (SEQ ID NO: 1).

- 59 -

55.   The protein of claim 54, wherein said protein is conformationally-constrained.

56.   A protein comprising the sequence Met-Val-Val-Ala-Ala-Glu-Ala-Val-Arg-Thr-Val-Leu-Leu-Ala-Asp-Gly-
5   Gly-Asp-Val-Thr (SEQ ID NO: 2).

57.   The protein of claim 56, wherein said protein is conformationally-constrained.

58.   A protein comprising the sequence Pro-Asn-Trp-Pro-His-Gln-Leu-Arg-Val-Gly-Arg-Val-Leu-Trp-Glu-Arg-
10   Leu-Ser-Phe-Glu (SEQ ID NO: 3).

59.   The protein of claim 58, wherein said protein is conformationally-constrained.

60.   A protein comprising the sequence Ser-Val-Arg-Met-Arg-Tyr-Gly-Ile-Asp-Ala-Phe-Phe-Asp-Leu-Gly-Gly-
15   Leu-Leu-His-Gly (SEQ ID NO: 9).

61.   The protein of claim 60, wherein said protein is conformationally-constrained.

62.   A protein comprising the sequence Glu-Leu-Arg-His-Arg-Leu-Gly-Arg-Ala-Leu-Ser-Glu-Asp-Met-Val-Arg-
20   Gly-Leu-Ala-Trp-Gly-Pro-Thr-Ser-His-Cys-Ala-Thr-Val-Pro-Gly-Thr-Ser-Asp-Leu-Trp-Arg-Val-Ile-Arg-Phe-Leu (SEQ ID NO: 10).

63.   The protein of claim 62, wherein said protein is conformationally-constrained.

- 60 -

64. A protein comprising the sequence Tyr-Ser-Phe-Val-His-His-Gly-Phe-Phe-Asn-Phe-Arg-Val-Ser-Trp-Arg-Glu-Met-Leu-Ala (SEQ ID NO: 11).

65. The protein of claim 64, wherein said protein
5 is conformationally-constrained.

66. A protein comprising the sequence Gln-Val-Trp-Ser-Leu-Trp-Ala-Leu-Gly-Trp-Arg-Trp-Leu-Arg-Arg-Tyr-Gly-Trp-Asn-Met (SEQ ID NO: 12).

67. The protein of claim 66, wherein said protein
10 is conformationally-constrained.

68. A protein comprising the sequence Trp-Arg-Arg-Met-Glu-Leu-Asp-Ala-Glu-Ile-Arg-Trp-Val-Lys-Pro-Ile-Ser-Pro-Leu-Glu (SEQ ID NO: 13).

69. The protein of claim 68, wherein said protein
15 is conformationally-constrained.

70. A protein comprising the sequence Trp-Ala-Glu-Trp-Cys-Gly-Pro-Val-Cys-Ala-His-Gly-Ser-Arg-Ser-Leu-Thr-Leu-Leu-Thr-Lys-Tyr-His-Val-Ser-Phe-Leu-Gly-Pro-Cys-Lys-Met-Ile-Ala-Pro-Ile-Leu-Asp (SEQ ID NO: 17).

20   71. The protein of claim 70, wherein said protein is conformationally-constrained.

72. A protein comprising the sequence Leu-Val-Cys-Lys-Ser-Tyr-Arg-Leu-Asp-Trp-Glu-Ala-Gly-Ala-Leu-Phe-Arg-Ser-Leu-Phe (SEQ ID NO: 18).

25   73. The protein of claim 72, wherein said protein is conformationally-constrained.

- 61 -

74. A protein comprising the sequence Tyr-Arg-
Trp-Gln-Gln-Gly-Val-Val-Pro-Ser-Asn-Trp-Ala-Ser-Cys-Ser-
Phe-Arg-Cys-Gly (SEQ ID NO: 19).

75. The protein of claim 74, wherein said protein
5   is conformationally-constrained.

76. A protein comprising the sequence Ser-Ser-
Phe-Ser-Leu-Trp-Leu-Leu-Met-Val-Lys-Ser-Ile-Lys-Arg-Ala-
Ala-Trp-Glu-Leu-Gly-Pro-Ser-Ser-Ala-Trp-Asn-Thr-Ser-Gly-
Trp-Ala-Ser-Leu-Ala-Asp-Phe-Tyr (SEQ ID NO: 20).

10   77. The protein of claim 76, wherein said protein
is conformationally-constrained.

78. Substantially pure DNA encoding the protein
of claim 54.

79. Substantially pure DNA encoding the protein
.15   of claim 56.

80. Substantially pure DNA encoding the protein
of claim 58.

81. Substantially pure DNA encoding the protein
of claim 60.

20   82. Substantially pure DNA encoding the protein
of claim 62.

83. Substantially pure DNA encoding the protein
of claim 64.

84. Substantially pure DNA encoding the protein
25   of claim 66.

- 62 -

85.   Substantially pure DNA encoding the protein
of claim 68.

86.   Substantially pure DNA encoding the protein
of claim 70.

5   87.   Substantially pure DNA encoding the protein
of claim 72.

88.   Substantially pure DNA encoding the protein
of claim 74.

89.   Substantially pure DNA encoding the protein
10   of claim 76.

90.   A protein isolated by a method comprising:
          (a) providing a host cell which contains
          (i) a reporter gene operably linked to a DNA-
binding-protein recognition site; and
15   (ii) a fusion gene which expresses a fusion
protein, said fusion protein comprising a test protein
covalently bonded to a binding moiety which is capable of
specifically binding to said DNA-binding-protein
recognition site;
20   (b) introducing into said host cell a second
fusion gene which expresses a second fusion protein, said
second fusion protein comprising one of said population
of proteins covalently bonded to a gene activating moiety
and being conformationally-constrained; and
25   (c) measuring expression of said reporter
gene; and
          (d) isolating a protein based on its ability
to alter the expression of said reporter gene when
present in said second fusion protein.

- 63 -

91. An interactor protein isolated by a method
comprising:

(a) providing a reporter gene operably linked to a
DNA-binding-protein recognition site;

5 (b) providing a first fusion protein, said first
fusion protein comprising a first protein covalently
bonded to a binding moiety which is capable of
specifically binding to said DNA-binding-protein
recognition site;

10 (c) providing a second fusion protein, said second
fusion protein comprising a second protein covalently
bonded to a gene activating moiety and being
conformationally-constrained, said second protein being
capable of interacting with said first protein;

15 (d) contacting a candidate interactor protein with
said first protein or said second protein;

(e) measuring expression of said reporter gene;
and

(f) isolating an interactor protein based on its
20 ability to alter the expression of said reporter gene
when present with said first protein or said second
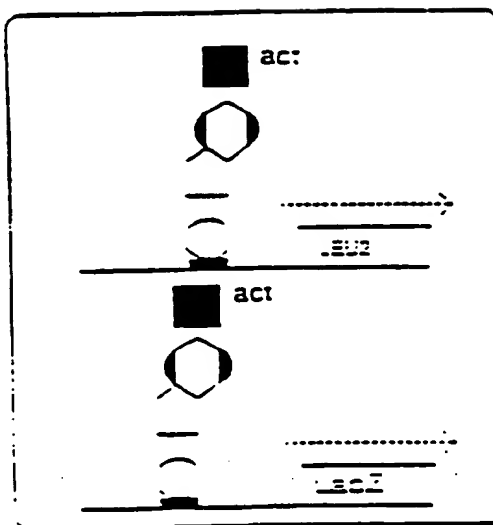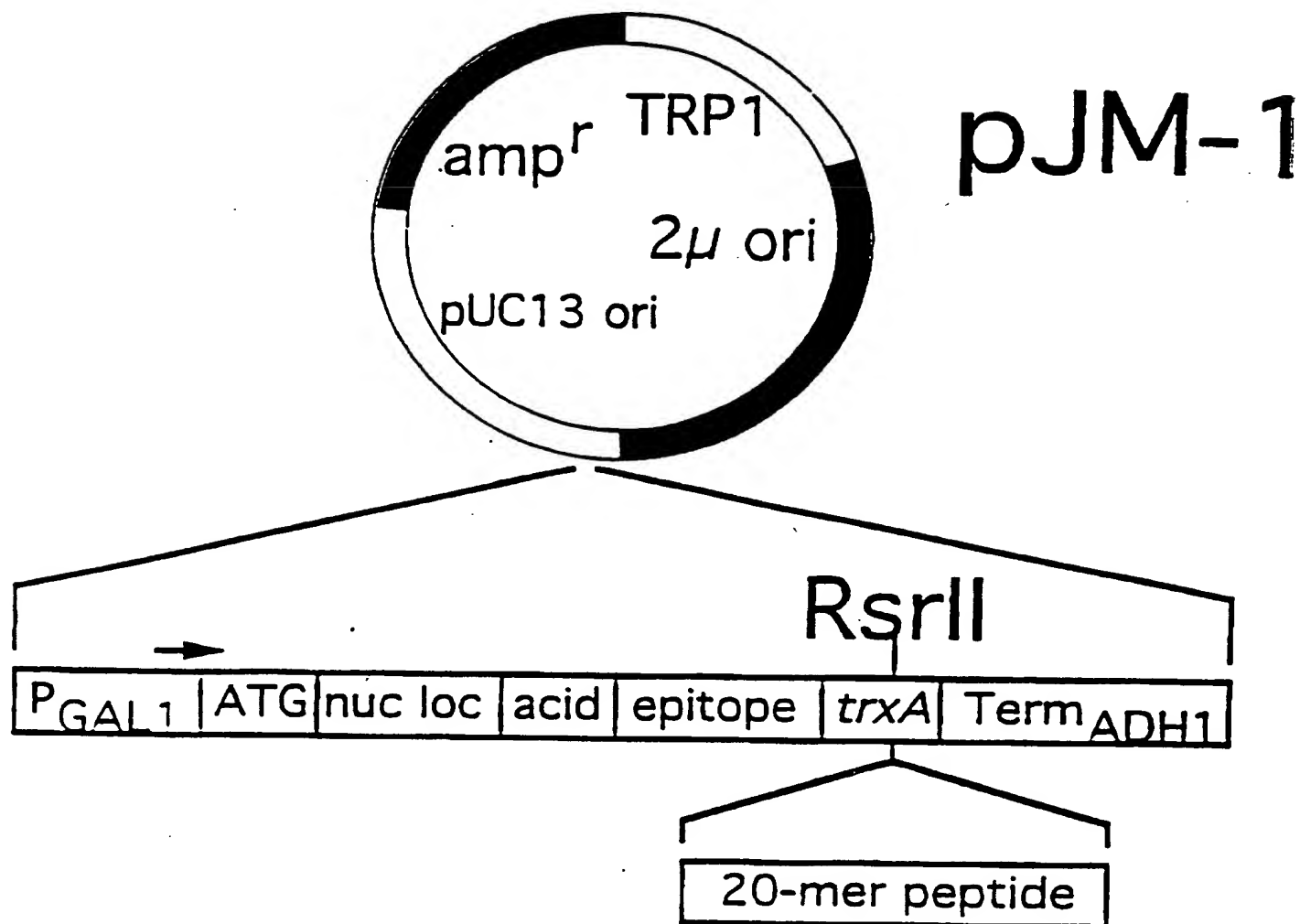protein.

FIG. 1 A
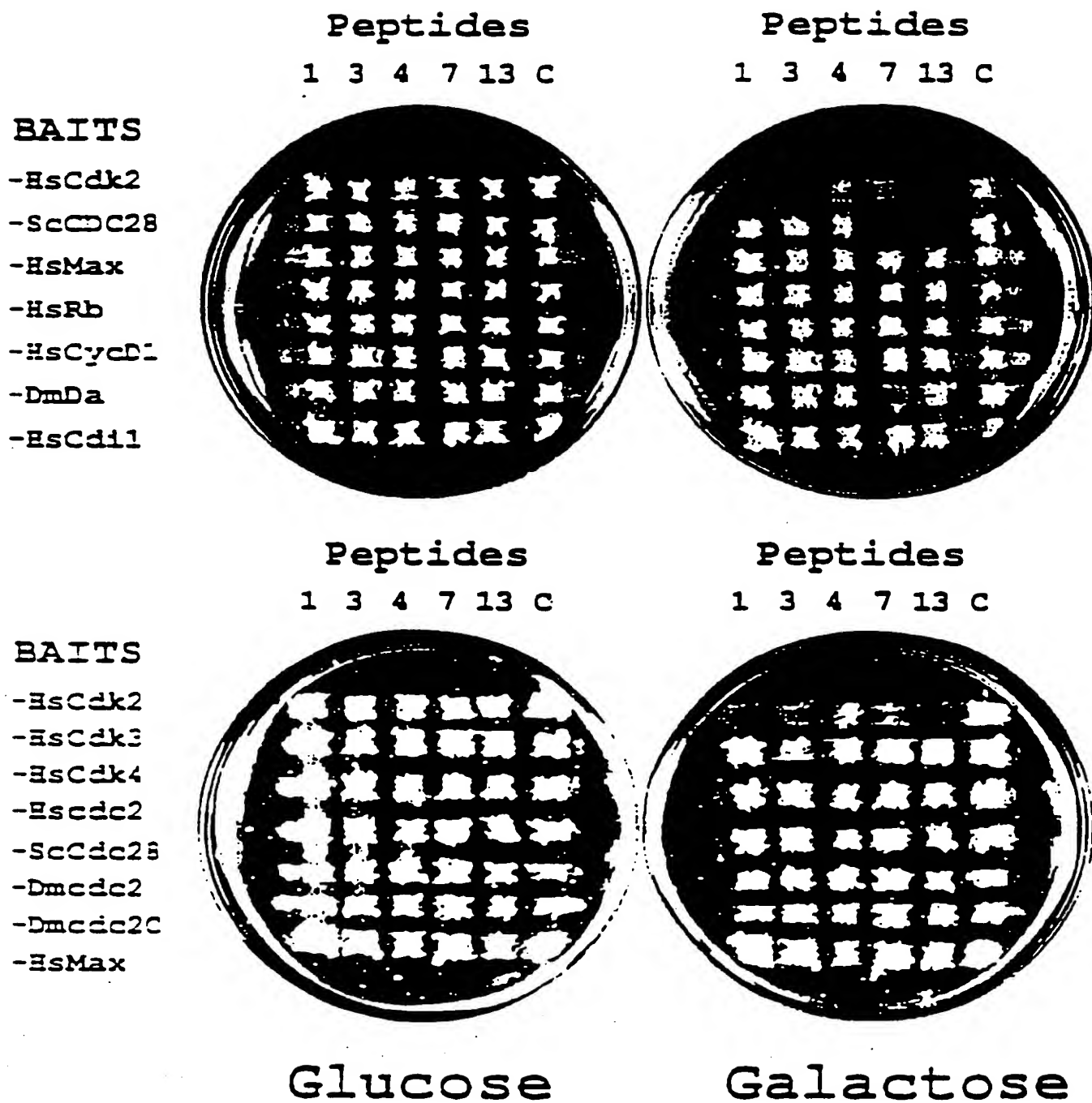
FIG. 1 B

FIG. 1 C

FIGURE 2

FIGURE 3A

Glucose   Galactose
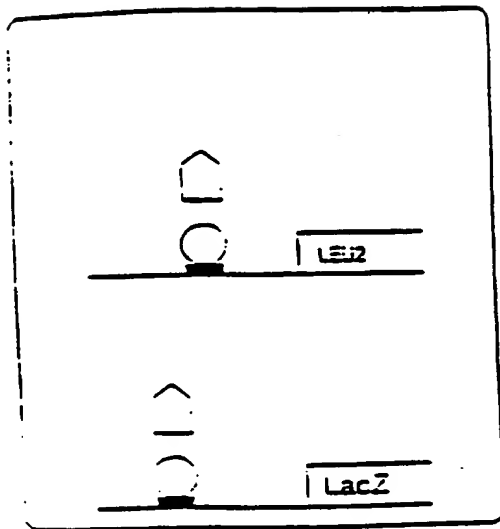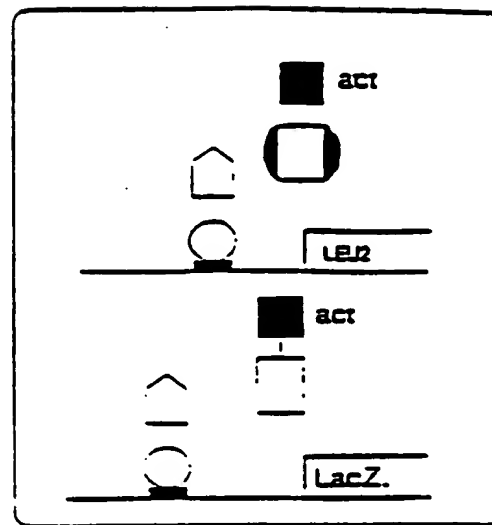
FIGURE 3B

Peptide 13: SVRMRYGIDAFFDLGGLLHG (SEQ ID NO:9)

Peptide 1: ELRHRLGRAL SEDMVRGLAW GPTSHCATVPG TSDLWRVIRFL (SEQ ID NO:10)

Peptide i5-1: YSFVHHGFFNFRVSWREMLA (SEQ ID NO:11)

Peptide i5-4: QVWSLWALGWRWLRRYGWNM (SEQ ID NO:12)
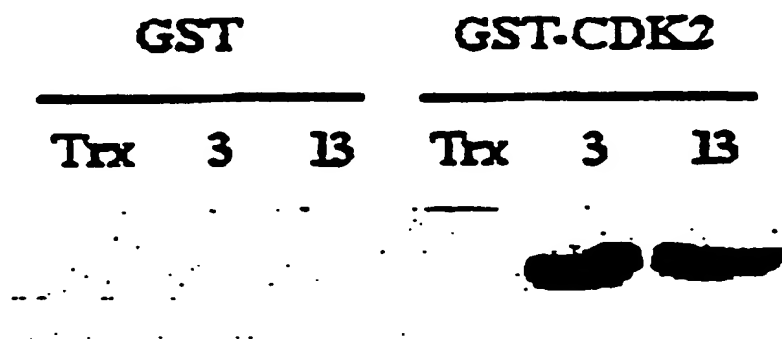
Peptide i5-9: WRRMELDAEIRWVKPISPLE (SEQ ID NO:13)

Peptide 3: LVCKSYRLDW EAGALFRSLF (SEQ ID NO:18)

Peptide 4: YRWQQGVVPS NWASCSFRCG (SEQ ID NO:19)

Peptide 7: SSFSLWLLMV KSIKRAAWEL GPSSAWNTSG WASLADFY (SEQ ID NO:20)

Figure 4

Figure 5
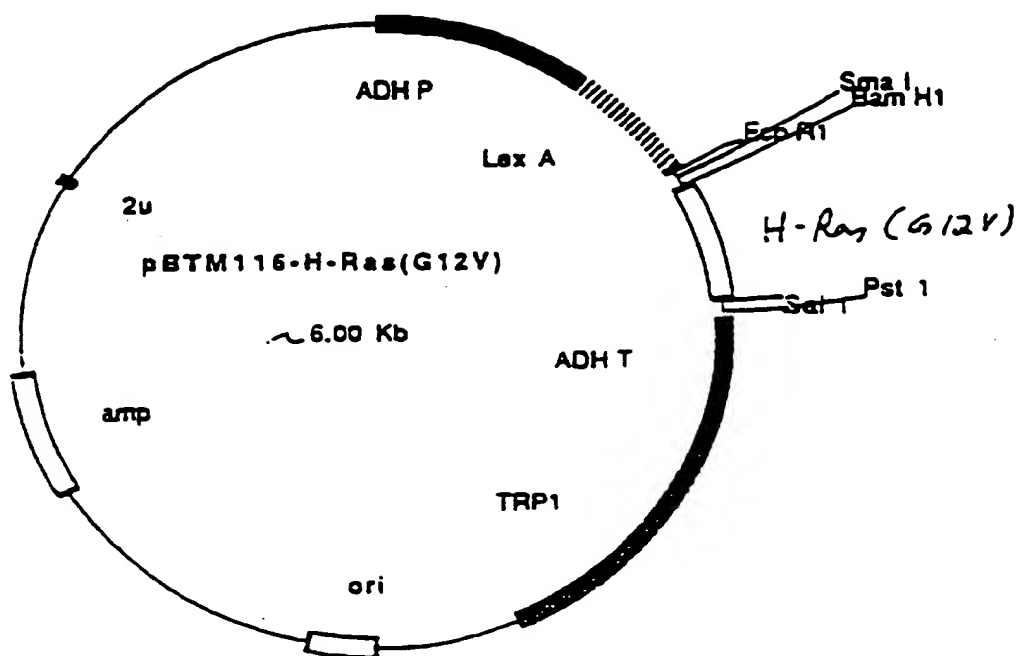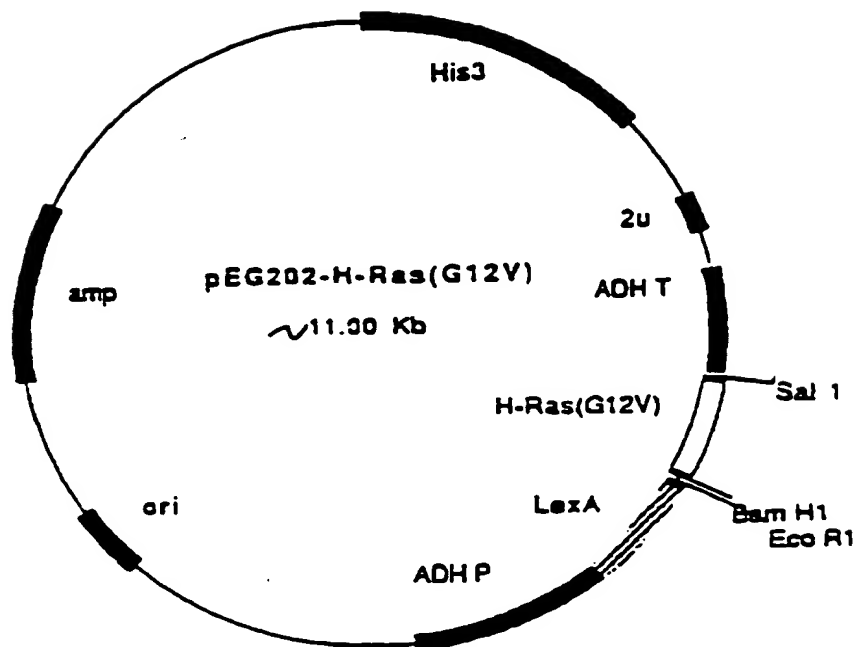
Figure 6

INTERNATIONAL SEARCH REPORT

A.	**CLASSIFICATION OF SUBJECT MATTER**

IPC(6)   :C07H 21/04; C07K 19/00; C12N 1/19, 5/10; C12Q 1/68; G01N 33/68
US CL   :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B.	**FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

  U.S. :   435/6, 7.1, 7.2, 7.21, 7.31, 240.2, 243, 254.21; 530/324, 326, 350; 536/23.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

  Please See Extra Sheet.

C.	**DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	Trends in Genetics, Vol. 10, No. 8, issued August 1994, Fields et al, "The Two-Hybrid System: an Assay for Protein-Protein Interactions", pages 286-292, see entire document.	1-53, 90-91
X --- Y	US, A, 5,283,173 (FIELDS ET AL) 01 February 1994, see entire document.	1-5, 13, 15-20, 27, 29-30, 44, 49-53, 90 --------------- 6-12, 14, 21-26, 28, 31-43, 45-48, 91

[X] Further documents are listed in the continuation of Box C.	[ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 OCTOBER 1995	**24 NOV 1995**
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer SCOTT D. PRIEBE Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	D. A. HARTLEY, "Cellular Interactions in Development, A Practical Approach" published 1994 by IRL Press (N.Y.), pages 153-179, see entire document.	1-5, 13-20, 27-30, 44, 49-53, 90 ------------- 6-12, 21-26, 31-43, 45-48, 91
X --- Y	Proceedings of the National Academy of Sciences USA, Vol. 88, issued November 1991, Chien et al, "The Two-Hybrid System: A Method to Identify and Clone Genes for Proteins that Interact with a Protein of Interest", pages 9578-9582, especially page 9578-9579.	1, 3, 5, 13, 15-17, 20, 27, 29-30, 44, 49-53, 90 ------------- 2, 4, 6-12, 14, 18-19, 21-26, 28, 31-43, 45-48, 91
Y	US, A, 5,270,181 (MCCOY ET AL) 14 December 1993, see columns 4-5, 10, and 12-16.	1-53, 90-91
Y	Journal of Biological Chemistry, Vol. 267, No. 25, issued 05 September 1992, Chakraborty et al, "Analysis of the Oligomerization of Myogenin and E2A Products *in Vivo* using a Two-Hybrid Assay System", pages 17498-17501, especially page 17500, paragraphs 2-3.	31-43, 91
Y	Cell, Vol. 75, issued 19 November 1993, Harper et al, "The p21 Cdk-Interacting Protein Cip1 is a Potent Inhibitor of G1 Cyclin-Dependent Kinases", pages 805-816, especially pages 805 and 806, column 2.	6, 25, 42

Form PCT/ISA/210 (continuation of second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/6, 7.1, 7.2, 7.21, 7.31, 240.2, 243, 254.21; 530/324, 326, 350; 536/23.4

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

USPAT, MEDLINE, BIOSIS, SCISEARCH, MPsearch_pp
search terms: interaction trap, two-hybrid system, protein#, protein-protein, peptide#, thioredoxin, DNA bind###, hybrid#, GAL4, LEXA, Fields S, Song O, Brent, R